

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: September 18, 2002, 10:35:58 ; Search time 54.03 Seconds  
(without alignments)  
4889.196 Million cell updates/sec

Title: US-09-695-795-4  
Perfect score: 7883  
Sequence: 1 MSIRLPHSIDRSASRQSHL.....EGSDIALDSATDVAVSPGP 1527

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 segs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

SPTRMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

VS 10/23/99

Result No.	Score	Query Match	Length	DB ID	Description
1	7883	100.0	1527	11 09ES67	09ES67 ratus norv
2	6482.5	82.2	1522	4 015085	015085 homo sapien
3	1822.5	23.1	1544	4 09NZNS	09NZNS homo sapien
4	1223	15.5	920	11 091VL3	091VL3 mus musculu
5	1216.5	15.4	919	11 061210	061210 mus musculu
6	1196.5	15.2	912	4 092888	092888 homo sapien
7	1180	15.0	919	11 092116	092116 ratus norv
8	1151	14.6	853	4 09BSB1	09BSB1 homo sapien
9	1139.5	14.5	869	4 000513	000513 homo sapien
10	1139	14.4	879	4 096F17	096F17 homo sapien
11	1022	13.0	750	4 015086	015086 homo sapien
12	962.5	12.2	540	4 096BF4	096BF4 homo sapien
13	944.5	12.0	2559	5 044381	044381 drosophila
14	936.5	11.9	2559	5 09V7X3	09V7X3 drosophila
15	922.5	11.7	2559	5 044113	044113 drosophila
16	672	8.5	979	5 019402	019402 caenorhabd1

17	556.5	7.1	2817	4 096P79	096P79 homo sapien
18	547.5	6.9	2813	4 096JP6	096JP6 homo sapien
19	531	6.7	1050	4 060274	060274 homo sapien
20	525.5	6.7	958	4 09H023	09H023 homo sapien
21	517.5	6.6	1429	4 09Y5T6	09Y5T6 homo sapien
22	492.5	6.2	985	11 09ESG7	09ESG7 mus musculu
23	486.5	6.2	985	11 0923E0	0923E0 mus musculu
24	457	5.8	893	4 09Y5T0	09Y5T0 homo sapien
25	451	5.7	930	4 096BS1	096BS1 homo sapien
26	449.5	5.7	164	11 088074	088074 mus musculu
27	442.5	5.6	864	4 094827	094827 homo sapien
28	346	4.4	1309	5 09V1V0	09V1V0 drosophila
29	343	4.4	1721	4 095216	095216 homo sapien
30	340	4.3	1714	11 0920R4	0920R4 mus musculu
31	340	4.3	1721	4 09UNK2	09UNK2 homo sapien
32	328	4.2	1676	4 09UGG4	09UGG4 homo sapien
33	328	4.2	1681	4 09NYG0	09NYG0 homo sapien
34	326	4.1	1696	4 09NZM3	09NZM3 homo sapien
35	310.5	3.9	790	4 09NVK9	09NVK9 homo sapien
36	301.5	3.8	1658	11 0920R6	0920R6 mus musculu
37	297.5	3.8	1737	5 09TW28	09TW28 dicystostell
38	297	3.8	1804	11 0921O1	0921O1 mesocricetu
39	296.5	3.8	1821	11 09R0C4	09R0C4 mus musculu
40	293.5	3.7	1234	5 09VZ85	09VZ85 drosophila
41	292.5	3.7	841	4 094989	094989 homo sapien
42	291.5	3.7	660	5 09NZV4	09NZV4 caenorhabd1
43	288.5	3.7	1065	11 091VS8	091VS8 mus musculu
44	288	3.7	2137	4 015021	015021 homo sapien
45	279.5	3.5	522	4 09H8F4	09H8F4 homo sapien

## ALIGNMENTS

RESULT 1  
ID 09ES67 PRELIMINARY; PRT; 1527 AA.  
AC 09ES67;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE RHOGEF GLUTAMATE TRANSPORT MODULATOR GTRAP48.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21133160; PubMed=11242047;  
RA Jackson M., Song W., Liu M.Y., Jin L., Dykes-Hoberg M., Lin C.-L.G.,  
RA Bowers W.J., Federoff H.J., Sternweis P.C., Rothstein J.D.;  
RT "Modulation of the neuronal glutamate transporter EAAT4 by two  
RT interacting proteins";  
RL Nature 410:89-93(2001).  
CC -i- SIMILARITY: CONTAINS 1 RGS DOMAIN.  
DR EMBL; AF225961; AAC28597.1; -;  
DR HSSP; Q12923; 3PDZ.  
DR InterPro: IPR001478; PDZ.  
DR InterPro: IPR001849; PH.  
DR InterPro: IPR000342; RGS.  
DR InterPro: IPR000219; RHOGEF.  
DR Pfam; PF00595; PDZ; 1.  
DR Pfam; PF00621; RHOGEF; 1.  
DR SMART; SM00228; PDZ; 1.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00315; RGS; 1.  
DR SMART; SM00325; RHOGEF; 1.  
DR PROSITE; PSS0106; PDZ; 1.  
DR PROSITE; PSS0003; PH\_DOMAIN; 1.  
DR PROSITE; PSS0132; RGS; 1.  
SQ SEQUENCE 1527 AA; 168533 MW; ABAEA20F41A3A9A CRC64;

Query Match 100.0%; Score 7883; DB 11; Length 1527;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRLPHSIDRSASKKOSHLSPISASMLSSLSLGDSTPERTSPSHRQPSDSETTAGL 60  
 DB 1 MSRLPHSIDRSASKKOSHLSPISASMLSSLSLGDSTPERTSPSHRQPSDSETTAGL 60

QY 61 VQRCVITIQKHQHGFGFTVSQDRIVLVQSVRPGGAMKAGYKGGDRITIKNGTWTNSSHL 120  
 DB 61 VQRCVITIQKHQHGFGFTVSQDRIVLVQSVRPGGAMKAGYKGGDRITIKNGTWTNSSHL 120

QY 121 EYVKLKSGAVYALTLTGSSPSPSVGSLQONSVAGVLRVNPITPPPPPPPLPPPOH 180  
 DB 121 EYVKLKSGAVYALTLTGSSPSPSVGSLQONSVAGVLRVNPITPPPPPPPLPPPOH 180

QY 181 TGKRPLODPEVOKHATQILMNMRLROEBEELQDILPPCGETSQRTCGRLSVDSQEADSGT 240  
 DB 181 TGKRPLODPEVOKHATQILMNMRLROEBEELQDILPPCGETSQRTCGRLSVDSQEADSGT 240

QY 241 DSGTERFPSTISESLMNRNYSVLSDPGLDSPQTSPIVILARAQHRRGSDAALLPLNHQGI 300  
 DB 241 DSGTERFPSTISESLMNRNYSVLSDPGLDSPQTSPIVILARAQHRRGSDAALLPLNHQGI 300

QY 301 DQSPKPLIIGPEEDYDPGVFNNSDIIIFODLEKSHPAVLVYELRYILSQADPGPLAFY 360  
 DB 301 DQSPKPLIIGPEEDYDPGVFNNSDIIIFODLEKSHPAVLVYELRYILSQADPGPLAFY 360

QY 361 LCSEVYQQTNPKDSRLGKDINWIFLEKNAPLRVKIPEMLOAEIDLRLRNEDPRNVLC 420  
 DB 361 LCSEVYQQTNPKDSRLGKDINWIFLEKNAPLRVKIPEMLOAEIDLRLRNEDPRNVLC 420

QY 421 AQAVALLEIOEQINDYRSKRTGLGSLYGENDLGLDGPRLRQMAEQALAGIILSK 480  
 DB 421 AQAVALLEIOEQINDYRSKRTGLGSLYGENDLGLDGPRLRQMAEQALAGIILSK 480

QY 481 YEEDRSAPMDFAVNTFMSHAGIRLRESRSSCTAEKTQSAAPDKKMWLPFFPKTKQSSNSK 540  
 DB 481 YEEDRSAPMDFAVNTFMSHAGIRLRESRSSCTAEKTQSAAPDKKMWLPFFPKTKQSSNSK 540

QY 541 KEKDALEDKKRNPLIRYIGKPKSSOSIKPGVNRNITIOHFENSHOYDVEPGTQRLSTGS 600  
 DB 541 KEKDALEDKKRNPLIRYIGKPKSSOSIKPGVNRNITIOHFENSHOYDVEPGTQRLSTGS 600

QY 601 FPEDLLESOSRSREIRLGRSGSLKGREEMKRSKKAENVPRPSDVMDAAAEARLHQSA 660  
 DB 601 FPEDLLESOSRSREIRLGRSGSLKGREEMKRSKKAENVPRPSDVMDAAAEARLHQSA 660

QY 661 SSSASSLSTRLENPTPPPTPKMGRRSIESPNLGFCTDYLPHLEDDLQGLSDLEPEPE 720  
 DB 661 SSSASSLSTRLENPTPPPTPKMGRRSIESPNLGFCTDYLPHLEDDLQGLSDLEPEPE 720

QY 721 VQNMQTVTKQDVANLTQREIDROEYINELFYTEASHLTKRLVDLIFVQRMKEMLMR 780  
 DB 721 VQNMQTVTKQDVANLTQREIDROEYINELFYTEASHLTKRLVDLIFVQRMKEMLMR 780

QY 781 BELARLFPMPLIELITHNSMCEAMKLRREGPIITRIDISPMKLARFGPRAEELQOVAAP 840  
 DB 781 BELARLFPMPLIELITHNSMCEAMKLRREGPIITRIDISPMKLARFGPRAEELQOVAAP 840

QY 841 CSYQVALELIRTKQKESRFQLFQOEASHPOCRRLQLRDLIVSEMRLLTKYPLLENI 900  
 DB 841 CSYQVALELIRTKQKESRFQLFQOEASHPOCRRLQLRDLIVSEMRLLTKYPLLENI 900

QY 901 IKHTBEGTSEHEKLCRADQCREILKRVNDAVKQTENRRLRLEGYQKRLDATALEASNP 960  
 DB 901 IKHTBEGTSEHEKLCRADQCREILKRVNDAVKQTENRRLRLEGYQKRLDATALEASNP 960

QY 961 AAEFKSLDTTKRMHIEGULTWRISSDKTLDLQVLLLEDLVVLQOEERLLLKSKSKA 1020  
 DB 961 AAEFKSLDTTKRMHIEGULTWRISSDKTLDLQVLLLEDLVVLQOEERLLLKSKSKA 1020

QY 1021 VCSSDSKQTFSPVLKNAVLIRSVATDKRAFFICTSELGPPQIYELVALTSSDKNIME 1080

DB 1021 VCSSDSKQTFSPVLKNAVLIRSVATDKRAFFICTSELGPPQIYELVALTSSDKNIME 1080

QY 1081 LLEEAQONATKHPGAPIRIHPSPGQEPAYOGSTSSRVEINDSEVYHTEKEPKKLPBG 1140  
 DB 1081 LLEEAQONATKHPGAPIRIHPSPGQEPAYOGSTSSRVEINDSEVYHTEKEPKKLPBG 1140

QY 1141 PGPEQVQOKLIAGEPVOEEDDEELRTLPRAPSPSLDENNGIRTRDYLALGPILM 1200  
 DB 1141 PGPEQVQOKLIAGEPVOEEDDEELRTLPRAPSPSLDENNGIRTRDYLALGPILM 1200

QY 1201 EGLADALDEVNRLILMSLLPGHTVKTQAAGEEDDLTPPSVSTISHMPDPSPG 1260  
 DB 1201 EGLADALDEVNRLILMSLLPGHTVKTQAAGEEDDLTPPSVSTISHMPDPSPG 1260

QY 1261 QAPTISDSTRLARPEGSGQEGEDVAVSSLAHLPPRTSSGVWDSPELDNRPAEAASTEP 1320  
 DB 1261 QAPTISDSTRLARPEGSGQEGEDVAVSSLAHLPPRTSSGVWDSPELDNRPAEAASTEP 1320

QY 1321 AASYKVRKVSLLPGGVAAKVAAGSNATPDSGQSFSELEVEGGAQATGNCFYVSMGAC 1380  
 DB 1321 AASYKVRKVSLLPGGVAAKVAAGSNATPDSGQSFSELEVEGGAQATGNCFYVSMGAC 1380

QY 1381 PLDSSTERTPTSPSPQCHSLPAMPTEPQYRGVNGQCSSLVRRDVYIFHTIQLTIK 1440  
 DB 1381 PLDSSTERTPTSPSPQCHSLPAMPTEPQYRGVNGQCSSLVRRDVYIFHTIQLTIK 1440

QY 1441 LHRKLMELAHRELKSLGESSGGTTPVGSFHTTEAARTDYSLSPPAKEALASQNGQ 1500  
 DB 1441 LHRKLMELAHRELKSLGESSGGTTPVGSFHTTEAARTDYSLSPPAKEALASQNGQ 1500

QY 1501 EOGSCPEBGSDIALBDSATDTAVSPGP 1527  
 DB 1501 EOGSCPEBGSDIALBDSATDTAVSPGP 1527

RESULT 2  
 015085 PRELIMINARY; PRT: 1522 AA.

AC 015085;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
 DE KIA0380 PROTEIN.  
 GN ARHGEF11/KIA0380.  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE=97344984; PubMed=9205841;  
 RA Tanaka T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,  
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. VII.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 4:141-150(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN;  
 RX MEDLINE=99150378; PubMed=10026210;  
 RA Fukuhara S., Mura C., Zohar M., Iguchi T., Gutkind J.S.;  
 RT "A novel PDZ domain containing guanine nucleotide exchange factor  
 RT links heterotrimeric G proteins to Rho.";  
 RL J. Biol. Chem. 274:5868-5879(1999).  
 CC -I- SIMILARITY: CONTAINS 1 RGS DOMAIN.  
 DR EMBL: AB002378; BAA20834.1; -;  
 DR HSBP: Q12923; 3PZ;  
 DR InterPro: IPR001478; PDZ.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR000342; RGS.

DR InterPro: IPR000219; RhogEF.  
 DR Pfam: PF00595; PDZ; 1.  
 DR Pfam: PF00621; RhogEF; 1.  
 DR SMART: SM00228; PDZ; 1.  
 DR SMART: SM00233; PH; 1.  
 DR SMART: SM00315; RGS; 1.  
 DR SMART: SM00325; RhogEF; 1.  
 DR PROSITE: PS50106; PDZ; 1.  
 DR PROSITE: PS50003; PH DOMAIN; 1.  
 DR PROSITE: PS50132; RGS; 1.  
 SO SEQUENCE 1522 AA; 167703 MW; CA16E125B9F8A4AA CRC64;

Query Match 82.2%; Score 6482.5; DB 4; Length 1522;  
 Best Local Similarity 82.3%; Pred. No. 0;  
 Matches 1278; Conservative 89; Mismatches 129; Indels 57; Gaps 10;

QY 1 MSIRLPHSIDRSASKKSHSLSPISAWLSLSLSDSTPERTSPSHHROPDSFTSETTAGL 60  
 DB 1 MSYLPQSIDR-----LSLSLSLSDSAPERKSPSHHROPDSASETT--GL 43  
 QY 61 VORCVIIQKOHGRTVSGDRIVLVQSVPRGGAAMKAGKEGRILIKVGTWYTNSSHL 120  
 DB 44 VORCVIIQKOHGRTVSGDRIVLVQSVPRGGAAMKAGKEGRILIKVGTWYTNSSHL 103  
 QY 121 EYVKLIKSGAYAAATLLGSSPSVSGVGLQONPSVAGVLRVNPRIIPPPPPPLPPQHI 180  
 DB 104 EYVKLIKSGAYAAATLLGSSPSVSGVGLQONPSVAGVLRVNPRIIPPPPPPLPPQHI 163  
 QY 181 TGPKPLQDPEVQKHATQILNMLRQEBELDILPPCGETSORTCEGLRVSQSEADSG 240  
 DB 164 TGPKPLQDPEVQKHATQILNMLRQEBELDILPPCGETSORTCEGLRVSQSEADSG 223  
 QY 241 DSGTERPSSISESMANNVSLSDPGLDSPOTSPYITLARVAQHRRQSDAALLPLNQGI 300  
 DB 224 DSGTERPSSISESMANNVSLSDPGLDSPOTSPYITLARVAQHRRQSDAALLPLNQGI 283  
 QY 301 DQSPKPLIIPPEEDYDPCYFNNESSDIFQDLKLSHPATLVYLRILISQADGPPLFY 360  
 DB 284 DQSPKPLIIPPEEDYDPCYFNNESSDIFQDLKLSHPATLVYLRILISQADGPPLFY 343  
 QY 361 LCSEVYQOTNPKDSRSIGKDIWIFLEKNAPLRVYKIPMLQAEIDLRLNNEDEPNVLC 420  
 DB 344 LCSEVYQOTNPKDSRSIGKDIWIFLEKNAPLRVYKIPMLQAEIDLRLNNEDEPNVLC 403  
 QY 421 AOEAVMLEIOEINDYRSKRTLGSLYGENDLGLDGDPLREKQAEKOLAAGDILSK 480  
 DB 404 AOEAVMLEIOEINDYRSKRTLGSLYGENDLGLDGDPLREKQAEKOLAAGDILSK 463  
 QY 481 YEEDRSAPMDFAVTFEMSHAGIRLRESRSSCTAEKTSAPDKDKWLPFPPTKXSSNSK 540  
 DB 464 YEEDRSAPMDFAVTFEMSHAGIRLRESRSSCTAEKTSAPDKDKWLPFPPTKXSSNSK 522  
 QY 541 KERDALEDKKRNPLIRYIGPKSSSSQ-----IKQGNVNIQIHFENSHQYDPE 590  
 DB 523 KERDALEDKKRNPLIRYIGPKSSSSQ-----IKQGNVNIQIHFENSHQYDPE 582  
 QY 591 PGQORLSTGSPFDLSDSSRSIRLRGSGSLKGREEMKRSKAEVPPRPSVDMAA 650  
 DB 583 PGQORLSTGSPFDLSDSSRSIRLRGSGSLKGREEMKRSKAEVPPRPSVDMAA 642  
 QY 651 AEAARLHQSSASSSLSSTRSLRNPPPTPKMGORSIESPNGCFIDVLLPHLLEDLG 710  
 DB 643 AEAARLHQSSASSSLSSTRSLRNPPPTPKMGORSIESPNGCFIDVLLPHLLEDLG 702  
 QY 711 QLSDLPEPEVQWQHTVGKDVANLQREIDRQEVINELFVTEASHLRTLRVLDLIFYQ 770  
 DB 703 QLSDLPEPEVQWQHTVGKDVANLQREIDRQEVINELFVTEASHLRTLRVLDLIFYQ 762  
 QY 771 RMKRENAMPREELARLPNLPPELLEIHNKSCAMKRTREGPIITROISDPMLARFGDPAR 830  
 DB 763 RMKRENAMPREELARLPNLPPELLEIHNKSCAMKRTREGPIITROISDPMLARFGDPAR 822

QY 831 EELQVAAQCFQSYVALTELIRTKORKESHQFQMEAEBSHPOCRRLQDLRLIVSEMQR 890  
 DB 823 EELQVAAQCFQSYVALTELIRTKORKESHQFQMEAEBSHPOCRRLQDLRLIVSEMQR 882  
 QY 891 TKYPLLENIITKTEGGTSEHEKLCRARDOCREILKRVNAVKOTENNRHLEGYQKRLA 950  
 DB 883 TKYPLLENIITKTEGGTSEHEKLCRARDOCREILKRVNAVKOTENNRHLEGYQKRLA 942  
 QY 951 TALERASNPLAEKSLDLTRKMHIEGRLTWMTISDKDITDLOVLLLEDLVLLQOEER 1010  
 DB 943 TALERASNPLAEKSLDLTRKMHIEGRLTWMTISDKDITDLOVLLLEDLVLLQOEER 1002  
 QY 1011 LLLKCHSKTAVGSSDSKOTFSPVLKLNVAVLRSVATDKRAFFITCISELGPQIYELVAL 1070  
 DB 1003 LLLKCHSKTAVGSSDSKOTFSPVLKLNVAVLRSVATDKRAFFITCISELGPQIYELVAL 1062  
 QY 1071 TSSDKNIMELLEBANONATKHPQAPIRPHSPGSGQEPAYQGSYSRVEINDSEYHAT 1130  
 DB 1063 TSSDKNIMELLEBANONATKHPQAPIRPHSPGSGQEPAYQGSYSRVEINDSEYHAT 1122  
 QY 1131 EKEPKLPEGPGPEQRYQDKQLAAGEPVQ--EDEBELRTLPRAPSLDGENRGIRTRD 1188  
 DB 1123 EKEPKLPEGPGPEQRYQDKQLAAGEPVQ--EDEBELRTLPRAPSLDGENRGIRTRD 1182  
 QY 1189 PVLATLTPGLMEGLADALEVDENLRHLILMGLPCHTVTKTOAAGEPEDDLPTPSVYS 1248  
 DB 1183 PVLATLTPGLMEGLADALEVDENLRHLILMGLPCHTVTKTOAAGEPEDDLPTPSVYS 1242  
 QY 1249 ITSHPMWPGSGQAFT--ISDSTRLARPEGSGEGEDVAVSLAHLPPRTRSSGVNDSP 1306  
 DB 1243 ITSHPMWPGSGQAFT--ISDSTRLARPEGSGEGEDVAVSLAHLPPRTRSSGVNDSP 1302  
 QY 1307 LDRNPAEAESTERPAASYKVRKVSLLPGGCGAAYAGSN---AIPDSQSSSELSEVE 1363  
 DB 1303 LDRNPAEAESTERPAASYKVRKVSLLPGGCGAAYAGSN---AIPDSQSSSELSEVE 1351  
 QY 1364 GGAQATGNCFYVMPAGPLDLSSTEGTPSPSPQCHSLPAMPTEPOPYRGVRGQGS--- 1420  
 DB 1352 GGAQATGNCFYVMPAGPLDLSSTEGTPSPSPQCHSLPAMPTEPOPYRGVRGQGS--- 1409  
 QY 1421 -----SLVRDQDVYFHTTEQLTIKHLRLKDMELAHRELKSLGSGSSGCTTPVCSFHT 1474  
 DB 1410 -----SLVRDQDVYFHTTEQLTIKHLRLKDMELAHRELKSLGSGSSGCTTPVCSFHT 1469  
 QY 1475 EAARWTDYSLSPAKKALASDSONGDEGSCPEGSDIALEDATPTAVSPGP 1527  
 DB 1470 EAARWTDYSLSPAKKALASDSONGDEGSCPEGSDIALEDATPTAVSPGP 1522

RESULT 3  
 O9NZNS PRELIMINARY; PRT; 1544 AA.  
 AC O9NZNS;  
 DT 01-OCT-2000 (TReMBLrel. 15, Created)  
 DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)  
 DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)  
 DE GUANINE NUCLEOTIDE EXCHANGE FACTOR.  
 GN Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PROSTATE;  
 RX MEDLINE=20160919; PubMed=10681437;  
 RA Kourlas P.J., Strout M.P., Becknell B., Veronese M.L., Croce C.M.,  
 RA Theil K.S., Krahe R., Runtu T., Knuttila S., Bloomfield C.D.,  
 RA Caligiuri M.A.;  
 RT "Identification of a gene at 11q23 encoding a guanine nucleotide  
 RT exchange factor: Evidence for its fusion with MLL in acute myeloid  
 RT leukemia.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 97:2145-2150(2000).





```

QY 580 FENSHOYDVPBGTORLSTGSPFEDLLEDSSSRSLRLGSGSLKREEMKSRKAENP 639
Db 280 -----PCRHUK-----VEADE-----KCPADRRKGGLGMSNRRTG 312
QY 640 RPRSDVMDMAAFAARLHQSASSASSLSTRLENPTPP-FPPKMRRIESPNIQFCYD 698
Db 313 TFGQD-----NGVSLHLSTDSVDSREP-GVDTPOBGTDPGQPTSLP----- 357
QY 699 VILPILLEDLQOLS-----DLEPEYVNMQHTYGVKDVANLTQREIDR 743
Db 358 -LAPPESTEDNGETESPEPEGDGDEPGRSGLELEPE-EPEGMRELVPDPTLLSPKQVQR 415
QY 744 QEYNLEFTEASHLTLVLVDLIFQMRKREMLPRELARLPPLLEIHNMCW- 802
Db 416 QEYISLTLTEAHVHMLVLDLIFQPMADGFFPLDELQNTFSLDELIEVHSLFLDR 475
QY 803 AMKKLEEGPIRLDISDPMILARFDGPARBELQOVAOFCYSVALERTKQKRSRQ 862
Db 476 LMKRRQESYLIEIGDVLARFDGAGSMFOKISSRFGSROSFALQJAKQKRRRC 535
QY 863 LEMQEAESHPOCRQLQRLDVLIVSEMORTKYPILLENIKHTEGTSEHEKLCRADQCR 922
Db 536 AFVQEAESHPCRCRLQKLMIPTEMQRLTKYPLLOSIGQNTSEET-ERKVELAECCR 594
QY 923 ELIKFPAENAVKOTENHRELEGYOKRLDATALERASNPILAEKSLDITRKMIHEGPTW 982
Db 595 ELIHNHNAVVRMEDLRLKDYQRLDTHLQSSSDPMLSEFRKNDITRKKLVHEBPLTW 654
QY 983 RISKDKTLQVLLLEDVLVLLQROBERLLKCHSKTAAGSSDSKQTFSPVLKNAVLR 1042
Db 655 RYTKKDAIEVHVLDDLLLLQRODERLLKSHSRTLPTPDGKMLRPLRLTSAMR 714
QY 1043 SVATDKRAFFITCSELGPQIYELVALTSSDKNIMWELLEAVONATKHPAIPRHP 1102
Db 715 EVATDHKAYVIFTWD-OEAQIYELVAOTSSERKNMCNLTETAGSLKYPADASRLKPR 773
QY 1103 SPGSGEAPVQOSTSSRVINSEVYHTEKEPKLPEG-----PPPE-----QRQ 1148
Db 774 SPSSIREPLL-----SSSENGTGAEMAPADARTERLNDLLPCRCPEQLAATALQKVL 830
QY 1149 D-KOLIAGEPVOEDEEELRTLPRAPSLDGEN-RGIRTRDPV-LTALGTPLMEGLAD 1205
Db 831 SLKQILLS-----TEEDS-----GAGPRDDGVPGRGAPVHNGQELIENLLSLEVAI 879
QY 1206 AALDEVE-----NLRLILMSLLPGHTVTKQAAGE 1235
Db 880 RQLELEEEFCRLRPL--SQLGTLSPMLAPE 911

```

## RESULT 6

```

Q92888 PRELIMINARY: PRT: 912 AA.
AC Q92888;
DT 01-FEB-1997 (TREMBlrel. 02, Created)
DT 01-FEB-1997 (TREMBlrel. 02, Last sequence update)
DE 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GUANINE NUCLEOTIDE EXCHANGE FACTOR P115-RHOGEF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN 111
RP SEQUENCE FROM N.A.
RX MEDLINE=96411762; PubMed=8810315;
RA Hart M.J., Sharma S., el Masry N., Qiu R.G., McCabe P., Polakis P.,
RA Bolla G.;
RT "Identification of a novel guanine nucleotide exchange factor for the
RT rho GTPase."
RL J. Biol. Chem. 271:25452-25458(1996).
DR EMBL: U64105; AAB17896.1; -.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RHOGEF.
DR InterPro: IPR003489; Ribosomal_S30.

```

```

DR Pfam: PF00621; RHOGEF; 1.
DR SMART: SM00235; PH; 1.
DR SMART: SM00325; RHOGEF; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
SQ SEQUENCE 912 AA; 102516 MW; 54AE5ACAFC291FIC CRC64;

Query Match 15.2%; Score 1196.5; DB 4; Length 912;
Best Local Similarity 34.2%; Pred. No. 2,3e-66;
Matches 335; Conservative 167; Mismatches 324; Indels 153; Gaps 32;

QY 308 IIGPE-EDYDPCYFNN--ESDIIIFODLEKTSHPYLVVLFYIISODPGLFLYLCSE 364
Db 23 IIGADEDEENLENTSEFQNSOFSLQEVKKRPALHVALQVLDQEPFLCLCHAD 82
QY 365 VYQOTNPKDSRLGDWNIWFLEKNAPLVKIPEMIAQETD--LRLNNEPRLVLCSEA 421
Db 83 MGLSGIPKRAKAFIDFVHSLFKETAVALRVPPVAVAFELDRADLLSEDVQRFFQVEY 142
QY 422 QEAVMLEIOEQINDYRSKRTGLGSLYGENDLGID-----GDPLREKMAEKOLAA 473
Db 143 VOSQVAVAGROLEDPRSKRMGMTP--WEQELAQLEAVWGRDASYEAREHVAERLLMH 200
QY 474 LGDI--LSKYEEDRSAPMDFAVTFEMSHAGIRLRSRSSCTAEKTOQAPDKDKMLPFF-- 529
Db 201 LEMQHTTISTDEKSAAYVNAIGLYRHLYGVTKSG-----DKSGRNFPRK 247
QY 530 -----PKTKROSSKKEKDALEDKKRNPIRLYIGPKSSSOSIKRQVNRITQ 578
Db 248 KYMGNRRSDPPKTKKGLSSI--LDAARWNNGEPQVDFRLKAEVDAEKQ----- 297
QY 579 HFENSHQIDVPEPGTQRLSTGSPFEDLLEDSSSRSEIRLGRSGSLKREEMKSRKAENV 638
Db 298 -----AYDRKGVGM-----SRD-----RNIGAPGDTPGVSLHPLSL 331
QY 639 RPRSDVMDMAAFAARLHQSASSASSLSTRLENPTPPFPKMRGISPNQFCYD 698
Db 332 DSPDEPGADAPL-----GDSFPGPMSLSLAPPESTDGAET-ESPEPG----- 378
QY 699 VILPILLEDLQ--LSLDEPEP-EVQNMQHTYGVKDVANLTQREIDROEYINLEFTEA 755
Db 379 -----DEGEPGRSGLELEPEPPGMRELVPDTHSLPKSQVQOEYISLVLTEA 429
QY 756 SHLRLRVLDLIFYMRKKNIMPREELARLPNLPPELIEHNSKCE-AMKKLEEGPII 814
Db 430 AAVRLRLVLDLTFQPMACELFPLELQNTFPSLDELIEVHSLFLDRIMKRROESGLI 489
QY 815 RDISPMLARFDGPARBELQOVAOFCYSVALERTKQKRSRQFQPMQEAESHPOC 874
Db 490 EETGVLARFDGAGSMFOKISSRFGSROSFALQJAKQKRRPCAFVQEAESHPRC 549
QY 875 RRLQRLDIVSEMORTKYPILLENIKHTEGTSEHEKLCRADQREILFYNEAVKQ 934
Db 550 RRLQKDMIPFEMQRLTKYPLLOSIGQNTSEPT-EREKVELAECCREILHVNQAVRD 608
QY 935 TENRRLEGYOKRLDATYALERSANPLAEFSLDTTKMKIHGEPYTRISKDITLDQY 994
Db 609 MEDLRDLQYORRDLSHLROSSDPMLEFPKNDITRKKLVHEGFLWVRVYKDAVEYHV 668
QY 995 LILEDLVLLQROBERLLKCHSKTAAGSSDSKQTFSPVLKNAVLRISVATDRAPFI 1054
Db 669 LILDDLLLLQRODERLLKSHSRTLPTPDGKMLRVLALTSAMTEVATDKRAFYVL 728
QY 1055 CTSSELGPQIYELVALTSSDKNIMWELLEAVON-ATKHPAARPIHPSPGSGEAPYQ 1113
Db 729 FTWD-QEAOIYELVAQVYSEKKNMCALITETAGSLKYPADASRPRP-RPRSSIREPLLS 786
QY 1114 GSTSSRVINSEVYHTEKEPKLPEG-----PGEPEYVOD-----KOLIAGEP 1158
Db 787 SSENGN--GGRETSPPARTERILSDLLPCRCPEQLAATALRKVLSLKLFLF--P 840
QY 1159 VOEDEEELRTLPRAPSLDGENGIRTRDPVLLALTGPLMEGLAD-----AALDEVE- 1212

```



Db 841 AEEDN-----GAGPRDGD--GVPGGLSPARTOE-IQENLLSEETMKOLEELE 889  
 Oy 1213 ---NLRLILMSILPGHTV 1228  
 Db 890 EPCRLRPL--SOLGAGNSV 906

RESULT 7  
 ID 092116 PRELIMINARY; PRT; 919 AA.  
 AC 092116;  
 DT 01-MAY-1999 (TEMBLrel. 10, Created)  
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
 DE LSC PROTEIN.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus;  
 OC NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Steinbremer H., Wuensche C., Seisler J.;  
 RT "Cloning of a rat homologue of Lsc, a mouse oncogene with structural  
 RT similarities to the Dbl family of guanine nucleotide exchange  
 RT factors";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ36911; CA15426.1; -  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR00219; RhGEF.  
 DR Pfam: PF00621; RhGEF.1.  
 DR SMART: SM00233; PH; 1.  
 DR SMART: SM00325; RhGEF.1.  
 DR PROSITE: PS50003; PH DOMAIN; 1.  
 SQ SEQUENCE 919 AA; 102597 MW; CF328CD2587717F8 CRC64;

Query Match 15.0%; Score 1180; DB 11; Length 919;  
 Best Local Similarity 33.2%; Pred. No. 2.5e-65;  
 Matches 334; Conservative 154; Mismatches 309; Indels 210; Gaps 35;

Oy 307 LITGPE-EDVDGPGFNESD--IIFODLEKKSHPATLVFLKRTIQADPGPLFYICS 363  
 Db 20 IITGAEDEDENELEAPEDONSOFOSEKRRPAHMLQVVALQFEFGPLCCLHA 79  
 Oy 364 EYVQOTNPKDSRSIGKIMINIFLEKNAPLRVKIPEMLOAEIDLR--LRNNEEDRVNICE 420  
 Db 80 DMLSSLPGEKKKAFDLDFYHSFLEKTVLKVPPVSAFELDRTRPDLISEDVORFTIQE 139  
 Oy 421 AQEAVMLEIOEQINDYRSKRTLGSLYGENDL---GLD--GDPLERQMAEQALAL 474  
 Db 140 VVGQQAQAAVTRQLEDFRSKRLMGMTPEQELSLLEPWIGDRGTYEARERIVARLLSHL 199  
 Oy 475 GDI--LSKYEEDRSAPMDFAVNTFMSHAGIRLRERSSCTAETQSAPEDKWLPEF--- 529  
 Db 200 EEMOHTISTDEKSAAVVTVTISLVRHLGVYTKSG-----DKKSGHNPFRKK 246  
 Oy 530 -----PKTKKSSNSKKEKDALEDKKRNPLRYIGKPKSSSOSIKPNVNRNIQH 579  
 Db 247 VMGKRKSDPEPKTKKGS-----IIDPARWNR-----GSPSAPD-----CRH 284  
 Oy 580 FENSHQYDVE-PCTORLSTGSPEDLLIESDSSREIRLGRSGSLAKREMKRSKAENV 638  
 Db 285 L-----KVEVDEKPG-----PAD-----RKGL-----GISSRRRTV 311  
 Oy 639 PRPSRDVMDAAEAARLHOSASSASSLSTR--SLENPPP--FTPMGRRSIESPULGF 695  
 Db 312 GTPQGD-----NPGVSLH---PLSYVSLDSREGVDTPOPGGTPTPOGFTSLP----- 357  
 Oy 696 CTVDILPHLEDDLGL-----SDLEPEP--EVQNMQHTYQKGVAVNLTQREI 741  
 Db 358 ----LAPPESTEDNGETESPEPGDGDGPRGSRGSLQEPDEEPGKRELVPDTLGLGPRNOV 413

Oy 742 DROEVINLEFVTEASHLRTTLDLIFQYRKREKNLMPRELTARLPNLPETIEHNSWC 801  
 Db 414 KRQEVISELVTETAAHVMRLVLDLTYQPAEGGFFLEQLNIFPSLDELVEHSLFL 473  
 Oy 802 E-AMKKLREEGPIIRIDISDPMARFDGPARRELOOVAOFCSYOSVALLETIRTKORESR 860  
 Db 474 DRLMKRROESGYLIEIGDVLARFDGAEQSWFQKISSRCSQSFALQEKAKQKREPR 533  
 Oy 861 FQLMQAEASHPOCRRLQRLIYSEMQRLTYPLLENIKTHTEGTSSEHEKLCRARQ 920  
 Db 534 FCAFVOEASERPRCRRLQKMDIPEMQRLTYPLLOSIGQNT-EPAEKAKVELAAEC 592  
 Oy 921 CRELLKPVNEAVKQETNRHREGYQKRLDAPALERASNPAAEKSIDLTRKMIHSGPL 980  
 Db 593 CRELLHYNQAVRKMEDILRLKDYQKRLDTHLQSDNPMSEFKNLDITKRLVHGSL 652  
 Oy 981 TWIRISKRTIDLOVLLLEDVLLQROEERLLKCHSKRTAVGSSSKQTESPVLKNAV 1040  
 Db 653 TWRLTKRAVEVHVLDDLLDILLQRODEGCSSHTSRTLTPTPDGKTMRLRVLRTSAM 712  
 Oy 1041 IRSVATDKRAFFICTSELAPQIYELVALTSSDKNIMELLEVAVON-ATKHPGAAPIP 1099  
 Db 713 TREVATDHKAFFVLTWD-GEAQIYELVAQTSERKSCALITETAGSLKVPAPASRPKP 771  
 Oy 1100 IHPSPGSEAPAYOGSTSSRVEINDESVYHTEKEPKKLPGEPGEQGVQKQILAAGEPV 1159  
 Db 772 -RPPSSRTRELLSSSEN-----GTGG-----TEAAPA 798  
 Oy 1160 QEEDEBELR-TLPAPPSLDGENRGITRPDVLATLGPMLMEGLADALEVDENLRHL 1218  
 Db 799 DARTERILINDLPEPCRGPEQO-----LAATFALQKVLSTKQIL 836  
 Oy 1219 LMSLLPGHTVKTQAAGEPD-DLTPTPSVSITSHPMDPGSPGQAPF 1264  
 Db 837 L-----STEEDSGAGPRPDGVP-----GGGAPGFTHT 865

RESULT 8  
 ID 09BSB1 PRELIMINARY; PRT; 853 AA.  
 AC 09BSB1;  
 DT 01-JUN-2001 (TEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TEMBLrel. 17, Last sequence update)  
 DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)  
 DE SIMILAR TO RHO GUANINE NUCLEOTIDE EXCHANGE FACTOR (GEF) 1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=MUSCLE, RHABDOMYOSARCOMA;  
 RA Strausberg R.;  
 RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC005155; AA05155.1; -  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR00219; RhGEF.  
 DR InterPro: IPR003489; Rhosomal\_S30.  
 DR Pfam: PF00621; RhGEF.1.  
 DR SMART: SM00233; PH; 1.  
 DR SMART: SM00325; RhGEF.1.  
 DR PROSITE: PS50003; PH DOMAIN; 1.  
 SQ SEQUENCE 853 AA; 95941 MW; 9426FC3B4CFD706C CRC64;

Query Match 14.6%; Score 1151; DB 4; Length 853;  
 Best Local Similarity 34.2%; Pred. No. 1.5e-63;  
 Matches 320; Conservative 166; Mismatches 313; Indels 136; Gaps 32;

Oy 342 VVELRITLSQADGPIFLYCSEYVQOTNPKDSRSIGKIMINIFLEKNAPLRKTIEMQL 401  
 Db 1 MALLQHVALLQEPGPLCCLHADMLSLGPKKAKKAFDLDFYHSFLEKTVLKVPPVPPNVA 60

OY	402	A E I D - - - L R L N N D P R V L C E A O E A V M L E J O R I N D Y R K R F L G S T Y G E U D L L G D -	457
Db	61	F E L D R T R A D L I S E D V O R F R V Q V S O Q V A V G R O L E D F R K R L M G M T P - W E D E L A Q L E A	118
OY	458	- - - - - G D P L R E O M A E K O L A L A G D I - L S K Y E D R S A P M D F A V N T F M S H A G I R L R E S R	508
Db	119	W Y G R D R A S Y E A R E H R H V A E R L L M L E E M Q H I T S D E E K S A A V N A I G L Y M H I L G V R T K S G -	177
OY	509	S S C R A E K Q S A P D K D K M L P F P P R K K K S S K K E K A L E D K K N P L T R I - - - - G R P K S S	564
Db	178	- - - - - D K S G R N F - - R K K Y M G N R S D E P K T K G L S T L D A A R M N R G E P Q - -	221
OY	565	S O S I R K P G N V R N I I O H F E N - S H O Y D V P P R - T O R L S T G S P F E D L L E D S S R S E I T R L G R S G S	622
Db	222	- - - - - V P D F R H L K A E V D A E K P G A T D R K G V G M P - - - - - S R D - - - - - R N I G	256
OY	623	L K G R E E K R S K A A N V R P R S D V D M D A A A R L H O S A S S A S L S T R L E N P T P P T F K	682
Db	257	A P G D T P G V S L H P L S L S P R E G A D P L L - - - - - G D S S P G M S L E S I A P E S T D	308
OY	683	M G R S I E S P N L G F C T D V I L P H L L E D D A G Q - L S D L P E P - E V O N M O H T G K D V A N I T O R	739
Db	309	E G A E T - E S P E R G - - - - - D E G E G R G L E P E R P E R G M E L Y P P T L I N S L R S	354
OY	740	E I D R O E V I N E L E V T E A S H L T R L Y A D L I F Y O R K R E N E M P R E L A R F P N D E L L E I H N S	799
Db	355	Q V K R O E V I S E L L Y E A H V M L R Y L H D L F P O P A E C F L F E L I O N I F P S D L L E I V H S L	414
OY	800	W C E - A M K L R E E G I I I D I D S P M A R P D G A R E L O Q V A O F C S O V S A L E L I R T K O R E	858
Db	415	F L D L R K M R O S G L Y L E I G V L L A R D G A E G S M F O K I S S F C S R O S F L E O L K A N O R D	474
OY	859	S R F O L F M O E A S H P C R R L O R D L I V S E M O R L K Y P L L E N I I K H T G S T S E H E K L C R A R	918
Db	475	P R C A F V O E A S R P R C R L O K D I P T E M O R L T K Y P L L L O S I G O N E E P T - E R E X E L A	533
OY	919	D O C E I L L F V N E A K O T E N E H R L E G Y O R L D A L A L E R A S P L A E K S I D L T T R K K I N E G	978
Db	534	E C C E I L L H N N O A R M E D L L R L K D V O R R L D I S H L K O S S D P M L S E K N D I T K K L V H G S	593
OY	979	P L T R I S K D T L D Q V L L E D L V L L O R O E B L L C H S K T A V G S D S K O T S P V L K L A N	1038
Db	594	P L T R V I T K D A V E N H V L L D D L L L L O Q R O E R L L K S H S T L P P T D G K T M L A P V L R L S	653
OY	1039	V L I N S A V T D K R A F I I C T S E L G P P O I E V A L T S O K N I M E L L E B A V O N - A T K I P G A P	1097
Db	654	A M T E R V A I D H A A F V L T W P - O E A Q I E L A Q V S E R K M M C A L I T E T A G S L T V P A P A S R P	712
OY	1098	I P I R S P G S O E P V A Q C S T S R V E I N D S E V Y H T E K E P P K L P E G - - - - - P G E O R V O D -	1148
Db	713	K P - R P S S T R E P L L S S E N G N - - - - G G R E T S P A D A T E R L I S D L P F C R P G E G O L A T A	768
OY	1150	- - - - - K O L I A G E P V O E D E E E L R T L P R A P S L D G E N K G I T R D P V I L A T G P L M E G	1203
Db	769	L R K V L S I K O L L F - - P A E E D N - - - - - G A G P R P D G D - G V P G G G L S P A R T Q E - I O E N	814
OY	1203	L A D - - - - - A L E D Y E - - - - - N L R L I L M S L L P G I T V	1228
Db	815	L L S L E E T M K O L E E L E E F C R L R P L L - S O L G N S V	847
RESULT	9		
ID	000513	PRELIMINARY:	PRT: 869 AA.
AC	000513;		
DT	01-JUL-1997 (Tremblrel. 04, Created)		
DT	01-JUL-1997 (Tremblrel. 04, last sequence update)		
DT	01-DEC-2001 (Tremblrel. 19, last annotation update)		
DE	SUBI.5 PROTEIN.		
GN	SUBI.5.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		

[illegible]

```

Db 649 RLTSAMTREVATDHKARFVLTWD-QEAOIYELVAQIVSERKNNMSALITERAGSLKVPAP 707
QY 1094 GAAPILPHSPSGQEPAYOGSTSSRVEINDSEVYHTEKEPKKLPPEGPGEPQROVKOLI 1153
Db 708 ASRRKP-RPSPSSREPLSSSEN----- 730
QY 1154 AAGEPVOEDEBELRT-----LPRAPSLDGENGRIGTRDPVLLATPLMEGLADA 1206
Db 731 --GNGGETSPADKTRIRILSDLLPFCRPGREGO-----LAAT 766
QY 1207 ALDEDVENRLHLLMSLLPGHTVKTQAGEPED-DLPTPSVSTSHPMWPGSPGQAPTI 1265
Db 767 ALKRVLSLKOL---LPPAE--EDNGAGPPRDGDGVPGGG-----PRSPARTQEI 810
QY 1266 SDRRLARPEGSGQEPGEDVAVSSLAHLPPRTRSSGVWDSPDL 1307
Db 811 QEN--LRLLETMTKTEELE-EFECRLRLPILLSQGLTSLPSL 849

```

```

RESULT 10
096F17
ID 096F17 PRELIMINARY; PRT; 879 AA.
AC 096F17;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, last annotation update)
DE UNKNOWN (PROTEIN FOR MGC:19628).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MUSCLE, AND RHABDOMYOSARCOMA;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011726; AA011726.1; -.
SQ SEQUENCE 879 AA; 98767 MW; 1D0863A5D1A57C9B CRC64;

```

```

Query Match 14.4%; Score 1139; DB 4; Length 879;
Best Local Similarity 33.5%; Pred. No. 8.6e-63;
Matches 326; Conservative 165; Mismatches 309; Indels 172; Gaps 35;
QY 308 IIGGE-BDVPDGYFNN--EDDIFQDLEKLSHPAYLVFLRYLTSQADGPGILFYLCSE 364
Db 23 IIGAEDEDENLEETNESEONSOFOLEQVRRRAHMLALQHVAFQFEGPLV----- 76
QY 365 VYQQTNPDKRSLSKIDWNIIFLEKNAPLRVYKIPMLQAEID--LRLRNEDPRNVLCSEA 421
Db 77 -----LNVPPVNNVAFELDRTRADLISDYGQRFYQEV 109
QY 422 QEAVMLEIQOINDYRSKRRLGSLGYENDLGLD-----GDPLREOMAEKOLAA 473
Db 110 VOSQQVAVGQQLDEFRRSKRLMGMP--WEQELQLEAMVGRDRASVBARRHVAERLLMH 167
QY 474 LGGD--LSKKEEDRSAPMDFAVNTFMSHAGRLRESRSSCTAEKTSQAPRKDKMLPPFPK 531
Db 168 LEEQHTITSDDEKSAVVVAIGLYMRHLVRYTKSG-----DKKSGRNF-- 212
QY 532 TKROSSSKKEKALDEKKNRPILRYI---GKPKSSSOSIKPGVNNIIIOHFN--SHQY 586
Db 213 RKKYMGRRSDEPAKTKKGLSSILDAARMNGEPQ-----VPDRHKLKAEV 256
QY 587 DVEPEG--TQRLSTGSPEDLLESDSSSETRILGSGSLKGREEMKRSKRAKNVPRPSDV 645
Db 259 DAEPGATDRKGGVGMF-----SRD-----RNIGAPGQDPGVSLHPDLSLSPDREP 305
QY 646 DMOAAARLHQSASSASLSSTRLENPTPTPKMGRSRSTSPVLGCTDVIILPHL 705
Db 306 GADAPLEL-----GDSSPQGMPSLESILAPPESTDEGAET--ESDPG----- 345

```

```

QY 706 EDDLGO--LSDLEPER-EVONMOHTVGKDVYANLTQREIDROEYINELFYEASHLTTR 762
Db 346 --DEGERGRSGLELPERPEPPGMRVELYPPDILHPSKQVQRQVEYISLLVTEAHYMR 403
QY 763 VLDFIYQRMKKNEMREELARLPNLPLEIHNMSCE--AMKLTRECPITRIDSDPM 821
Db 404 VLHDLFPQMAECLEFPLEELQNIFFSLDELIEVHSLFLRLMKRRQESGLYLEIGDVL 463
QY 822 LARFDGARELDQVAAOFCSYSALELRTQQRKSRRQLTMOQEKSHPOCRRLQLD 881
Db 464 LARPDGEGSWFQKISRFCRSQSFALQKARQDPFCARVQEAESRPRKRRLQDK 523
QY 882 LIVSEMORLTKYPLLLENIITKTEGCTSEHEKICLRADOCREILKFVNEAVKOTENHRL 941
Db 524 MIPTEMORLTKYPLLLOSIGONTNEPT--EREKVELAECCREILHNHNVQAVRMEDLRL 582
QY 942 EGYOKRIDATALEBASNPPLAEFKSLDTLTRKMIHSGPLTWIRISKOTLDLOVLLDVL 1001
Db 583 KDYQRRLDLSHLROSSDPMLESEFNKNDITKKKLVHEGRLTWRYTKDAVEVHLLDDL 642
QY 1002 VILROBERLLTCHSKTAAGSSDSKOTFSPLKINAVLIRSVATDKRAFFIICTSBLP 1061
Db 643 LILQRODERLLKSHSRTLPPTPDGKTMRLPVLRLSAMTREVATDHKARFVLTWD-QE 701
QY 1062 POIYELVALTSDKNIMWELLEAVON--ATKHGAAPIPHPSPSGQEPAYOGSTSSRY 1120
Db 702 AQIYELVAQIVSERKNNCALITERAGSLKVPASRRKP-RPSPSSREPLSSSEN- 759
QY 1121 EINDSEVYHTEKEPKKLPPEG-----PGEPQROVD-----KQIIAAGEPVOEDEE 1165
Db 760 --GGERTSPADARERILSDLLPFCRPGREGQALATRLKRVLSLKOLLF--PAEDN- 812
QY 1166 ELRLTPRAPSLDGENGRIGTRDPVLLATPLMEGLAD-----AALEPVE-----NLRH 1216
Db 813 -----GAGPPRGD--GVGGGFLSPARTQE-IQENLLSLEMTKOLELEEFECRLP 863
QY 1217 LILMSLLPGHTV 1228
Db 864 LL--SQLGNSV 873

```

```

RESULT 11
ID 015086 PRELIMINARY; PRT; 750 AA.
AC 015086;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, last sequence update)
DE KIAA0382 PROTEIN (FRAGMENT).
GN KIAA0382.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA MEDLINE=9734984; PubMed=9205841;
RA Nagase T., Ishikawa K., Nakajima D., Ohira M., Seki N., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro."
RL DNA Res. 4:141-150(1997).
DR EMBL; AB002380; BAA20836.1; -.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000219; RhGEF.
DR Pfam; PF00621; RhGEF; 1.
DR SMART; SM00233; PH; 1.
DR SMART; SM00325; RhGEF; 1.
DR PROSITE; PSS0003; PH_DOMAIN; 1.
FT NON_TER 1
SQ SEQUENCE 750 AA; 84746 MW; 8B14AE449C26CE72 CRC64;

```

Query Match 13.0%; Score 1022; DB 4; Length 750;  
 Best Local Similarity 32.8%; Pred. No. 1.4e-55;  
 Matches 269; Conservative 132; Mismatches 232; Indels 188; Gaps 25;

QY 750 LEVTEASHLTLRLVLLIFQMRKKNLMPREELARLPNLPLELIHNSWCEAMKRLR 809  
 DB 1 LFYTERAHVHTLTKLVQVDFQVRVSRREGIILSPSELARKIFSLIEDIQLDHLGLNQMKAVKR 60  
 QY 810 --EGPIIRDSVPMARFPQPARRELOOVAOCYSQVALELIRKORRESFOLFMOE 867  
 DB 61 RNFTSYIDQIGEDLTLWFSFGPEEKLKHAATPCSNOPLALEMTKSRKKDSRFQFVD 120  
 QY 868 AESHPQCRRLQRLDLIVSEMQRLLTKYPLLENIKHTEGTSEHEKLCARADOCREILRF 927  
 DB 121 AESNPICRRRLQDLIPTQMQRLTKYPLLDNIATKYTEMT--EREKVKKAADHCQILNF 179  
 QY 928 VNEAVKOTERNRRLRLEGYOKRDLATALERASNPAAEFKSIDLTRKMIHEGPILTRISND 987  
 DB 180 VNOAVEAEKKORLEDYQRLDITSSLKSEYVNEELRNLDLTKRMHIEGPIVWKVND 239  
 QY 988 KTLDLVLLLEDVLLVLOROERLLKCHSKTAVGSSDKQSPVPLKINAVLIRSVATD 1047  
 DB 240 KTIDLTLLEDLVLVLOQDDRLVLRCHSKILASTADSKHTFSPIKISTYLVRQVAYD 299  
 QY 1048 KRAFLITCTSELGPPQIYELVALTSSDKNIMWELAEAVONATKHPGAAPRIHPSPG- 1106  
 DB 300 NKALFVISMDSNG-AQIYELVAQTVSEKTYWQDLICR--MAASVKEGTFKPIPLQSTPGE 357  
 QY 1107 -----SQEPAYQSTSSRVIND----- 1124  
 DB 358 GDNDEEDPSKLEEOHGISVTGLQSPDRDLGLESTLISSKPOSHLSTSGKSEVRLFYA 417  
 QY 1125 ---SEVYTEKEPKKLPEGS-----PGEPQR-VOD-----KQLAAGEPVQE 1161  
 DB 418 ERQFAEQHTDGLTKVEGDYQAIIPDSHLPVSEEMALDALRNGLKQLLVQGLGLE 477  
 QY 1162 ED-EEELKRLPRAPSLDENNGCIRTRDPVLLALITGPLMEGLADALEVENVRLHLLM 1220  
 DB 478 KSVQEDWQHPR-----YRT-----ASQGP-----QTDVIONSENK--AY 512  
 QY 1221 SLPGHTVQTAAGEPDDLTTPPSVVSITSHMPDPSGQAPITISDTRL----- 1271  
 DB 513 HSEGGMPRTGDIATCYSPKSTESFA--PRD--SYGLAPQDSQASNIILVMDHMTF 568  
 QY 1272 ----ARPEGG--QPEGEDVAVSSLAHLPPR--TRSSGVWDSPELDNR----- 1310  
 DB 569 PEMPTMEPEGGLDSDGEHFDAREAHSDENPSEGDAVKNKEKDVNLRTISGNVLLDGYD 628  
 QY 1311 PAEAASTEPASAKYKRVVSLIPGGGVGAAYAGSNA--IPDSGQSESL-----E 1361  
 DB 629 PVOESTDEAVS-----SLTQPMGIPAVESTHQQHSPOHTSDGASIPTEPEFLVO 683  
 QY 1362 VEGGAQATNCFCVSNMAGPLDSTREPTGTPSPSQCHSLPAMPTEPPQYRGVGGQCSS 1421  
 DB 684 QRMGAAMEY--SCEPIQSPSSCADSQS----- 708  
 QY 1422 LVRRDVIYFHTLEQLTKRLKMLKMELAHRELKSLGSES 1462  
 DB 709 -----IMEXYHKTEADLEHLKKEVESYTIICORLASA 741

RESULT 12  
 096BF4 AC PRELIMINARY; PRT; 540 AA.  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DE HYPOTHEICAL 61.3 KDA PROTEIN (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=PRIMARY B-CELLS FROM TONSTILS;  
 RA Strausberg R.;  
 RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC015652; AAH15652.1;  
 KW Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 540 AA; 61278 MW; BD25E9E5662DE009 CRC64;

Query Match 12.2%; Score 962.5; DB 4; Length 540;  
 Best Local Similarity 42.0%; Pred. No. 4.6e-52;  
 Matches 227; Conservative 105; Mismatches 160; Indels 49; Gaps 16;

QY 714 DLEPEVQNMQHTGKDVAVNLQREIDROVINELPVTEASHLTLRLVLLIFQMR 773  
 DB 17 ELEPE-EPPGMRELVPDDTLHSLPSQVAKQVEYISLTVTAHAYMLVHLDFQFPA 75  
 QY 774 KENLMPREELARLPNLPLELIHNSWCE--AMKKLREEGPIRDISDPMARFQPARRE 832  
 DB 76 ECLFPELELQNIFFSLDELIVHSLFLDRMLKRRQESYLEELGVDVLAFFDAGESW 135  
 QY 833 LQQAQAQCSYQVALELIRTKORRESRQFLMOEAESHPOCRRLQRLDIYSEMQRLLK 892  
 DB 136 FOKISRSFCSRSQFALQOLKAKORKDPRCAFAVOEAESRRCRRQLKMIPTEMQRLTK 195  
 QY 893 YPLLENIKHTEGGSEHEKLCARADOCREILKTVNEAVKOTERNRRLRLEGYOKRDLATA 952  
 DB 196 YPLLQSIGQNTPEEP--EREKVELAECCRETLHNVQAVRMEDLRLKDYQRLDUSH 254  
 QY 953 LERASNPAAEFKSIDLTRKMIHEGPILTRISKDTLDVLLLEDVLLVLOROERLL 1012  
 DB 255 LQSSDPMSEFKNLDITKKLVLHEGPIWVTKQXAVENVHLLDLDLRLQDERLL 314  
 QY 1013 LKCHSTAVGSSDKQSPVPLKINAVLIRSVATKRAFFITCTSELGPPQIYELVALTS 1072  
 DB 315 LKSHSKTLLPTPDGKTMRLPVLRITLSAMTREVATDHKAAYVLTWD--QEAQIYELVAQTV 373  
 QY 1073 SDKNIMWELAEAVON-ATKHPGAAPRIHPSPQSOEPAVQYSTSSRREINDSEVYTE 1131  
 DB 374 SERKMWALITFTAGSLKVPAPASRKP--RSPSSSTRELLSSSNGN---GGRITSAD 429  
 QY 1132 KEPKKLPEG-----PGEQVQVOD-----KQLAAGEPVQDEBELTLPAPPS 1176  
 DB 430 ARTERILSDILPFCRPPGEGQLAATALKRVLSLKQLLF---PAERDN-----GAGPP 478  
 QY 1177 LDGENRGITRDPVLLALITGPLMEGLAD-----AALDEVE-----NLRLHILWSLPGHT 1227  
 DB 479 RQGD--GVPGGGLSPARTQF--IQENLISLETMKQLEELPEFCRLRPL--SOLGNS 533  
 QY 1228 V 1228  
 DB 534 V 534

RESULT 13  
 044381 AC PRELIMINARY; PRT; 2559 AA.  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DE SHAR PEI/DRHOGEF2.  
 GN RHOGEP2 OR CG9635.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxId=7227;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=98088790; PubMed=9428514;  
 RA Barrett K., Leptin M., Seltman J.;  
 RT "The Rho GTPase and a putative RhoGEF mediate a signaling pathway for  
 the cell shape changes in Drosophila gastrulation."  
 RL Cell 91:905-915(1997).  
 DR EMBL; AF032870; AAC38820.1; -.  
 DR HSSP; P29476; 10AV.  
 DR FLYbase; FBgn0023172; RhoGEF2.  
 DR InterPro: IPR001525; C5\_DNA\_meth.  
 DR InterPro: IPR002219; DAG\_PE-bind.  
 DR InterPro: IPR001478; PDZ.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR000219; RhoGEF.  
 DR Pfam; PF00130; DAG\_PE-bind; 1.  
 DR Pfam; PF00595; PDZ; 1.  
 DR Pfam; PF00169; PH; 1.  
 DR Pfam; PF00621; RhoGEF; 1.  
 DR SMART; SM00109; C1; 1.  
 DR SMART; SM00228; PDZ; 1.  
 DR SMART; SM00233; PH; 1.  
 DR SMART; SM00325; RhoGEF; 1.  
 DR PROSITE; PS00094; C5\_MTASE\_1; UNKNOWN\_1.  
 DR PROSITE; PS00479; DAG\_PE\_BIND\_DOM\_1; 1.  
 DR PROSITE; PS0081; DAG\_PE\_BIND\_DOM\_2; 1.  
 DR PROSITE; PS0106; PDZ; 1.  
 KM Phorbol-ester binding.  
 SQ SEQUENCE 2559 AA; 280857 MW; 112CCEFF3362BCDF CRC64;

Query Match 12.0%; Score 944.5; DB 5; Length 2559;  
 Best Local Similarity 20.9%; Pred. No. 6.9e-50;  
 Matches 482; Conservative 279; Mismatches 651; Indels 895; Gaps 82;

QY 17 QSHLSPIASWSSLSLSDSTPERTSPSHRPPSDTSETTAGLVQRC-----VII 67  
 DB 204 QSDNSNPVLOAPGERSLNLTPLSRDLSGHTQESTPTATPTPSLALPKNFQYLTLTV 263  
 QY 68 OKDOHGFVSGDRIYLVQSVRPGAGAKYKGGRIKVTNGTWTNSHLEVYKLIK 127  
 DB 264 KDSNKGKMGKVSQDNVFEVSVKPGCAAEIAGIAGDMILRVNGHEVRLKHPYVGLIK 323  
 QY 128 SGAYALATLIGSS-----PSPVGV-----SGLQONPSVAGVLRV----- 161  
 DB 324 ASTTVELAVKRSQKLTSPSSSVYTPSTPILSGRDTASITGPQVDSIKRREMETYKIQ 383  
 QY 162 -----NP----- 163  
 DB 384 TLQKMLEQKLNLERLKSQDNNSYKLSSEANIKRLREQLHQVGAEDAPTVKIQAAAGNKN 443  
 QY 164 --TIIP-----PP-----PP----- 170  
 DB 444 TALLTPNOIOLHSATSHNQGFHLLHHNNHNNNTPPOQOASTSPALSLPLRSLS 503  
 QY 171 ----- 170  
 DB 504 LSLGTRKNKTEKDLTSSPGLTDFLQOQRMGSHQAESMSQSMHQTSTPTSOQFFPHQ 563  
 QY 171 -----PPPLP-----PPQ-----HITG 162  
 DB 564 QOHRKETGPTSGKKNKFLISRLIEDVPPPLPORNPPQLNLDKNGNAGSGSHLVA 633  
 QY 183 P-----KPLDPEVKAHTOILMM 202  
 DB 624 PVSDDLRATSPOLNRSQOQDLPRSTDNSPNAKSKSKITKALSDP---KMSOTMLQ 680  
 QY 203 ----- 202  
 DB 681 ESASAGAAGSIEVDGPPPLPPLPGMTEDMSRGSCONLAQPNVSTAFNPLVSTT 740  
 QY 203 -----LRQEEELQIL-----PPGGEFSQR----- 223  
 DB 741 TAVQNDNLNIAFPLSORPNIVQQLQYQOQOQHOMSGGATGALGOTPNLGNKKNHRRVGS 800

QY 224 -----TCEGRLSVDSQEADGLDSTGERPSESILM-----NRNS 259  
 DB 801 SPDNMHRHPRDRTTKTSSGWEIYENDGESS-PPGTPPLPYLSSIMHTVLEDNENNRGA 859  
 QY 260 VLSDPG--LDSPO-----TSPVILARVQH--HRRQSDALLPLNHQIDOSPPLIT 309  
 DB 860 AAGPGVFIESHOFTPMAGASPIPLSHSSHMAAOSND-----TQKEIIS 906  
 QY 310 GREPDVD-----PCYFNESDIIIFODELKLKSHPAVLVFLRYILSOADGPLLFLYL 361  
 DB 907 MEDNSDLDEPFTIDENGPENNLRLLEAE-----NVTFAITLNVYISNDPAPLFLYL 960  
 QY 362 CSEVYOQTNPKDSRSLGKDIWNITFLEKNAPLR-VKIPEMLOAIFDLRLRNEDP-----RN 416  
 DB 961 ITFLYKEGTCKDMKKNAYELHSTFLVPRALSLWRODESLAREVDNVLQLEYKVELKRT 1020  
 QY 417 VLCEAOEAVMLEIQEOINDYRSKRTIGLSLYGEND--ILGLDGPLREKMAEKOL--- 471  
 DB 1021 VFLRSRRRAKDLISEQLREFQOKRTAGLGYIGPTDCLKAEAKTDKLR-QIIDKYLMPN 1079  
 QY 472 -----AALGDILSKYEEDRSAPDFA--VNTFMSH----- 499  
 DB 1080 LHALIEDENGSPEDYKVALCALSTVIYRIENTPRPSSIVERVHHFVSRDKSEKSRIT 1139  
 QY 500 -----AGIR-----LRE-----SRSSCTAEKRTQS 518  
 DB 1140 MGKNRKNVAGHPLVLRQYEVYHCHNCQRTIINGVSPQGYHCTDKKNIHROCSKYVDES 1199  
 QY 519 APDKKWLPFFPKTKKOSNSKKEK-----DALEDKRR----- 551  
 DB 1200 CPGP-----LPOAKRLAHNDKISKFMKIRPRTSDVIGNEKRSRODELDVELTPDRGQ 1253  
 QY 552 -----NPILRYIGKPKSSQSIKGNVNNIIQHENSQYVVPBG----- 592  
 DB 1254 ASTVROPDRRPANISIRKNGTSCNTSGL---NTTDLOSSFPGSCANDSINPGGAGC 1310  
 QY 593 TORLST-----GSFPE-----DLLESDSSR-----SEIRLGRS 620  
 DB 1311 NMDLSTSVASTPTSTSGSVNAGLSAFALNALDTYVKEAREREYQSPHEKSNAPVSNRS 1370  
 QY 621 GSLKGRREKRSKRAEN-----VPRPSVDVMDAAAEARLRHQASASSASLS----- 668  
 DB 1371 ESYKERLSNKRNRNRSRKRTDPSLSRP-NDEQDLGLSNATYVGSNSLSAGSGESP 1429  
 QY 669 TRSLSENTPP-----FTPMGRSSTESPNL-----GCTDYLPHLEDDLGQS-- 713  
 DB 1430 STSMENHPAAGAGVQVPPMGLNQNHPLHLIQHAAQYCCQDSSFQAGLAGAAGSSAAS 1489  
 QY 714 -----DLEPEEVO--NMQHTVCGKVANLTQREIDREIYINELFVT 753  
 DB 1490 NSSFWNAGHPLPARWTLSEDEDEDVNEADMSWVAEVALALTDMAKKQEIINELYOT 1549  
 QY 754 EASHLRTLRVLDLIFYORMKKNLMPREELARLP-PNLPLLEIHNSECAMKLR-EBG 811  
 DB 1550 ERNHHVRLTKLIDLFLPLLESGLISQDHLFLPPALSLREIHGAFESOLSKORREHN 1609  
 QY 812 PIIRDISDPMLARFPDGAARELQOVAOFCOSYSVALIELRTKORKSRRLOLMOEAESH 871  
 DB 1610 HVAVNTIGDLLADMFGDGGVGLCEFAQFCARQOIALEAKERNKDEMIQKLKLSSESH 1669  
 QY 872 POCRRLOLRDLIVSEMOURLKGYPLLENITKHT-----EGGSTSEHEKILCARDOCREILKF 927  
 DB 1670 KACRRLLEKRLPLTVLQRLKRYLPLENLKVVYRLPLEMTTEAEVIAQRAVESKRLIVE 1729  
 QY 928 VNEAVKOTENRRHRLLEGYQKRLDATALERASNPPLAAEFKSIDLTTRKMIHEGPLTRISKD 987  
 DB 1730 VNOAVKTAEDAKHLKIONIORLIDRSSYDK-----EEKCKIDLTOHNLIHGNGTL--IKKN 1781  
 QY 988 KTLDIQVLLLEDLVLLOROEERLLK-CHSKTAVGSSDSKOTFSPVLTKLNA-VLIRSV 1045  
 DB 1782 PSVOLGHELFENMTVLLTKODDKYLLKNLHTPLST-----TNKVPSPIMSIDADTLIRQEA 1837

```

OY 1046 TDKRAFFICTSELGPQIYELVALTSSDKNIMMELLEBAVONATKH----- 1092
DB 1838 AAKNSFFELI---KMTSOMLELRASSSECKTWFKEHSDVAAROSKNRKNASNHDTSI 1894
OY 1093 ---PGAAPPIHSPPGSOD----- 1118
DB 1895 SPPALALAP-HSNTYESLESLTDYVOPPLAATLTTPPLAPMLPATVTPPATATNNSNVS 1953
OY 1119 -----RVEINDSEVYHTEKEPKKLPGEPPREORVODKOLI-AOGEPOV-----EED 1163
DB 1954 SLTGVOLRNPORDATASESDADYVNTPKPRSSONEVYNTFRMSIRSGEPLOKXSANGTEAN 2013
OY 1164 EEEELR-----TLRPRAPSLDGENR-----GIRTRDPVLLAL-----TGLUMEG-LA 1204
DB 2014 DVTLLRHSOSTRESVRSRGSTGEERNSTYGVANGSKNSDASIVCSNNSNTRTLMOQSPV 2073
OY 1205 D-ALEDEVNLRHLLMLSLPGHFKVTOAGEPED-----DLTPRP-----SVASI 1249
DB 2074 DFTALQVSTISPAHNTAEVLTTPBEKRLRDLASTIRNDLEKOKITICIFILRPEHYQIYDI 2133
OY 1250 TSHPMWPGSPGQAPRTISDSTRLARPEGSG-----PEGEVAVSSL-----AH- 1291
DB 2134 AMMP-----EAPKOSADIALAAYDQIOTLTKMLNEYMHVPEQEVSAVSTAVGCHHE 2186
OY 1292 -----LPRTRSSGVWDSPELDNRNPAEAASTEPASTYVVKVSLPPGCGGAAYV 1343
DB 2187 KEKLRKKVAPSSSSSSPPPLPPRNQHOAOQIIPP-----RLMPRLQTLDDV----- 2238
OY 1344 AGSNAIPDSGQSESELSEVEGGAQATGNCFYVMPAGPLDSSTERTPTSPSCCHSLPA 1403
DB 2239 ---AHHEDDOCYCEIDELR-----LPAIPKPHERRP-PLAP----- 2272
OY 1404 WTEPOPIRGVGVGOCSSLVRRDVYIFHTT-----EQLTILHLRL- 1444
DB 2273 FNTPEKTSQSV---IDASKRQSTDAVPEGLLEOEPLGDKTETKGEDEVNVTVPSDKLS 2328
OY 1445 ---KDMELAHRELKSLGSESSGCTTPVGSFHTEAARWTDLSLSPAKELASPOSQO 1500
DB 2329 ESCNEREQCEVDITREKVAADPTTSKNEAASVDELPSQSRKIKTAENAKSVADKEDNE 2388
OY 1501 EOGSCPEEGSDIALEDSATNTAVSPCP 1527
DB 2389 E---TIEKGV-ASTVDSSTQTSPTESP 2411

RESULT 14
OYV7X3 PRELIMINARY: PRT: 2559 AA.
AC OYV7X3:
DT 01-MAY-2000 (TREMBLrel. 13. Created)
DT 01-MAY-2000 (TREMBLrel. 13. Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19. Last annotation update)
DE RHGEF2 PROTEIN.
GN RHGEF2 OR CG9635.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Dirosophillidae; Drosophilla.
OX NCBI_TaxID=7227;
RN
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abill J.F., Abgayan A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballev K.Y., Benos P.V., Berman B.P., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

```

```

RA Butris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Duthin K.J., Evangelista C.C., Ferraz C., Ferliera S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaser K.,
RA Gladis N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz C., Kulp D., Lai Z.,
RA Lasro P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon R., Nusser D.R., Pacle J.M.,
RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,
RA Splier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003804; AAF57915.1; -.
DR HSSP: P29476; 10AV.
DR FLYBASE: FBgn0023172; RhogEF2.
DR InterPro: IPR001525; C5_DNA_meth.
DR InterPro: IPR002219; DAG_PE_bind.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001849; PH.
DR InterPro: IPR002119; RhogEF.
DR Pfam: PF00130; DAG_PE_bind.1.
DR Pfam: PF00595; PDZ.1.
DR Pfam: PF00169; PH.1.
DR Pfam: PF00621; RhogEF.1.
DR SMART: SM00109; C1.1.
DR SMART: SM00228; PDZ.1.
DR SMART: SM00233; PH.1.
DR SMART: SM00325; RhogEF.1.
DR PROSITE: PS00094; C5_MTASE_1; UNKNOWN_1.
DR PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE: PS50106; PDZ.1.
SQ SEQUENCE 2559 AA; 280805 MW; 07725DE04982605C CRC64;

```

```

Query Match 11.9%; Score 936.5; DB 5; Length 2559;
Best Local Similarity 20.8%; Pred. No. 2.2e+49;
Matches 481; Conservative 279; Mismatches 652; Indels 895; Gaps 82;

OY 17 OSHLSPPASMLSSLSGDSPTPEKTPSHRQPSDTEETAGLVORC-----VII 67
DB 204 QSDNSNPVLOAPGERSSLNTPLPSRLDSGHTQESTPATTPSPSLAPKPNFOYLTLT 263
OY 68 OKDOHGFETVSGDRIVVQSVRPGAAKGVKEDRIIRKNGMTVNSSHLEVVKLIK 127
DB 264 RKDSNGVGMKYSGDNPNVVESEYKPGGAELAGLVAGDMILRVNGHEVLEKHPYVGLIK 323
OY 128 SCVAAALTLGLSS---PPSVGV-----SGLOQNPVSVAGVLRV----- 161
DB 324 ASTVVELAVKRSOKLTRSSVSVYTPSPILSGRDRFASITGPOVPDVKRRMETYKIQ 383
OY 162 -----NP----- 163
DB 384 TLQKMLEGEKINLEKLSKDONNPYSKISEANIRKRLQELHGVGAEDAPTVKLOAAAGNKN 443
OY 164 --IIPP-----PP----- 170

```

```

Db 444 TALLTPNOIHLASATSHNOQFHHHNNNNYPOOQOPASTPAPLSLPRSLSS 503
Qy 171 ----- 170
Db 504 LSLGTRKNTKTEKDLTTSSPFGELTTDFLQOORMSHOAESMSOSMOHTSTPSOQFFPHQ 563
Qy 171 -----PPLP--PPO-----HITG 182
Db 564 QOHRFKETGPTSKGNKFLSLRLIEDVPPPLPQRNRPQRLNDLKNNGASPGSHLVA 623
Qy 183 P-----KPLDPEVOKHATQILMM 202
Db 624 PVSDDLBRATSPOLNRSSOQOOLPSTSDNSPSNAKSKRSKIKTKALSDP--KMSYOMLOM 680
Qy 203 ----- 202
Db 681 ESASAGAAGSIEVGDGPPPLPRLPGMMTEDMSRGSCONLAOPNSVGAFTNPLYSTT 740
Qy 203 -----LROEEELQDIL-----PGGETSQR-- 223
Db 741 TAVONDMINIAFLPSQRPNIYQLOOYQOQOQOHOMSGQATGALGOTPNLGNKRRHVS 800
Qy 224 -----TCEGRLSVDSOAGDSGLDSGTERFBSISLM-----NRNS 259
Db 801 SPDMMHRAHPDRITKTSGSWEIYEKDGESS--PGTPEPPLSSHMTLEDPNENNRGA 859
Qy 260 VLSDPG--LDSPO-----TSPIYLARVAOH--HRROGSDAALLPLNHGIDGSPKPLI 309
Db 860 AAGPGVFIESHQFTPMAGASSPIPISLHSHMAAOSND-----TQKELIS 906
Qy 310 GPEEDVD-----PGYNNESDIIIFODLEKLSHPAYLVFLRYLSOAGPGLPYL 361
Db 907 MEENSLDEPFIDENGPFNLTFRLEAE-----NVTFLAIFLNYVISNDPAPLLFYL 960
Qy 362 CSEYQOOTNKDSLSLGDIMNTFLEKNAPLR--VKIREMLOAEIDLRLNNEDP-----RN 416
Db 961 ITELYKGTGSKDMKMAVEIHSFTLVPRAPLSWYRQDESLAREVDNVLQLEYDVEILRT 1020
Qy 417 VLCEAOEAVMLEIQEINDYRSKRTGLISLYGEND--LLGLDDPPLREROMAEKOL-- 471
Db 1021 VPLRSKRRAADLISEQUREFOQKTAGLITGYPTDKLAEAKDKLRF--QIIDKYLMPN 1079
Qy 472 -----AALGDILSKYEEDRSAPMDFA--VTENFSH----- 499
Db 1080 LHALIEDENGSPREDVARKVALCSALSTVIYIRIENFRPPSSIVERVHHFVSROKFSKRI 1139
Qy 500 -----AGIR--LRE-----SSSSCAEKTQS 518
Db 1140 MGKRRKNMNVGRHPLVLYQYEVTHCHNCQTTIMGVSPQYHCTDCKINHROCSKVDES 1199
Qy 519 APDKDKMLPEFPKTKKOSNSKKEK-----DALEDKR----- 551
Db 1200 CPGP-----LPQAKRLAHNDKISKFMGKIRPTSDVIGNKRSRQDELDVELTPRQ 1253
Qy 552 -----NPLRITIGPKSSOSITKPGNVRNIIIOHFNSHOYDVEPG----- 592
Db 1254 ASIYROPDRRPDANISIRSGNTSCNTSGL--NTTDLOSFFHGSCANDSINGGAGC 1310
Qy 593 TORLST-----GSFPE-----DLLESDDR-----SEIRLGRS 620
Db 1311 NMDLSTVASTTSTSGSVAAGLSAFALNALDLYDKEARRERISQHPKHSAYVSNRS 1370
Qy 621 GSLGREEMRSRKAEN-----VPRPRSDVMDAAEAARLQSASSASSSLS----- 668
Db 1371 ESKYERLSNKRNRNRSRKTSDPSLSNR--NDEQLDLGLSNATYGGSSNSLSAGSGESP 1429
Qy 669 TRSLIENPPP-----FTPKMGRSISPNL-----GCTDYILPHLELDIGOLS-- 713
Db 1430 STSMERHAPGAAGGVOPPMGLNQNPHLLIOHQAQOQOOSFOAGLAGAAGSSAAS 1489
Qy 714 -----DLEPEPEVO--NMQHTVGKDVANLQORELDROEVNLEFVT 753
Db 1490 NSSFWNAGHPLPVARMTLESEDEDDVNEADWSSNVAAEVALALDAEKKROEITINEIYQT 1549

```

```

Qy 754 EASHLRTLRVLDLIFYORMRKENIMPREELARLE--PNIPELIEIHNSEAMKRLR--EG 811
Db 1550 ERNHVRLTKLIDLRLFLPLVYESGLLSODHLLLEPPRALLSIRELHGAPEOSLKRRIEHN 1609
Qy 812 PIIRDIDPMLARPDPAREELQVAAOFCSYOSVAALELIRTKOKRESROLFMQAEASH 871
Db 1610 HVVNTIGDLADMDGOGSVLCEFAAOFCAQOIALEALKERKNDEMLOKILKSESH 1669
Qy 872 POCRRLODLIYSEMOQLTKYPLLENIITKH-----EGGISEHEKLCRAPDQCREILKF 927
Db 1670 KACRRLEKDLPLTVLQRLTKYPLLFENLYKVYRLLPENTTEAEALORAVESSKRIIVE 1729
Qy 928 VNEAVKOTENNRHLEGYOKRLDATALERASNPILAEPKSDLTTRKMIHEGPTWIRISK 987
Db 1730 VNQAVKTAEDAKHQNOROKRIDRSSYDK-----EEFKKDLITQHLIHGNLTL--IKN 1781
Qy 988 KTLDLQVLLLEDLVLLQROEBELLK--CHSKTAVGSSDSKQTESPYVLKUNA--VLIRSA 1045
Db 1782 PSVOLHGLEFENMIVLLTKODDKYLLKNLHTPLSI-----TKRPVSPIMSIDADTLIRQEA 1837
Qy 1046 TDKRAFFIICSELGPPQIYELVALTSSDKNIMMELLEAVONATKH----- 1092
Db 1838 ADKNSFFLI--KKTSOMLELRAPSSSECKTWFKHPSDVAAROSKNKSNASHDTSI 1894
Qy 1093 --PGAPPIPIHSPPGSOE-----PAYOGSTSS 1118
Db 1895 SDPLALAIIP--HSNKRSELESTDTVOPLAATLTTPLPLAPMLPIAVTTPAPANNVS 1953
Qy 1119 -----RVELNDESVYHTEKEPKKLPBGEPGEORVODKOLI--AOGEPVQ-----EED 1163
Db 1954 SLNGVOLRNPORDATASESDADYVNTPKPRSSQNEVKTMSIRSTGEBIDOKYSANGTEAN 2013
Qy 1164 EEEELR-----TLPRAPSLDGENR-----GIRTRDPVLLAL-----TGPLMEG--IA 1204
Db 2014 DVTLRHSQSTRESVRPGSTGEERNSTGYMGNGSKRDSATVCSNNSNNTFTLLMQSPLV 2073
Qy 1205 D-AALEDEVNRLILMSLPGHTVKTQAAGEPD-----DLPTTP-----SVVSI 1249
Db 2074 DPTAIQVSIAPHAHAEPVLTPGEXLRLDASIRNDLLEKOKIICODIFRLPVEHYDQIYDI 2133
Qy 1250 TSHWPDGSEGOAPTISDSTRLARPEGSQ-----PGEDEVANSLS-----AH- 1291
Db 2134 AMPP-----EAPKSDADIALAAVYDQIOTLTKMLNEMHVTPEQEVSAVSTAACGICHE 2186
Qy 1292 -----LPRTRSSGVWDSPELDNRPAEAASTEPASVYKVRKYSILPBGVGAAKY 1343
Db 2187 KEKLRKVAPESSFSSEPPPLPPPNROHQAQOAIIPS--RLMPKLOTLDLDEV----- 2238
Qy 1344 AGSNAIPDSGSESELSVEEGGAQATGNCIFYVSNPAGPLDSTEGTPSPSQCHSLPA 1403
Db 2239 --AIHEDDDGCEIDELR-----LPAIYSPKHEBPT--TPLAP----- 2272
Qy 1404 WPTPEQPIRGYRGQGCSSLVARDVYIFHTI-----EDLTIKLRL- 1444
Db 2273 FNTPEKTSQSV--IDASKQOSTDAVEGLELEQEPLEGDKTETKGEDNEVKYVPSDKLS 2328
Qy 1445 --KDMELARELLKSLGESSGCTTPVGSFHTHEAARWTDYSLSPAKKALASDSQNG 1500
Db 2329 ESCHEERQCYAADITKEVADPTTSKNEMAASVDELPPQSREIKTAENASKVAADKEDNE 2388
Qy 1501 EOGSCPEGSDIALEDSATDTAVSPG 1527
Db 2389 E--TIEBGV--ASTVDSSTQTSPTESP 2411

```

```

RESULT 15
O44113
ID O44113 PRELIMINARY; PRT; 2559 AA.
AC O44113;
DT 01-JUN-1998 (TEmblrel. 06, Created)
DT 01-JUN-1998 (TEmblrel. 06, last sequence update)
DT 01-DEC-2001 (TEmblrel. 19, last annotation update)

```

DE PUTATIVE GUANINE NUCLEOTIDE EXCHANGE FACTOR RHOGEF2.  
 GN RHOGEF2 OR CG9635.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC NCBI\_TaxId=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Haacker U., Perrimon N.;  
 RL Genes Dev. 0:0-0(1998).  
 DR EMBL: AF031930: AAB8816.1; -  
 DR HSSP: P29476; 1QAY. -  
 DR Flybase: FBgn0023172; RHOGEF2.  
 DR InterPro: IPR001525; C5\_DNA\_mech.  
 DR InterPro: IPR002219; DAG\_PE-bind.  
 DR InterPro: IPR001478; PDZ.  
 DR InterPro: IPR001649; PH.  
 DR InterPro: IPR000219; Rhogef.  
 DR Pfam: PF00130; DAG\_PE-bind; 1.  
 DR Pfam: PF00595; PDZ; 1.  
 DR Pfam: PF00169; PH; 1.  
 DR Pfam: PF00621; Rhogef; 1.  
 DR SMART: SM00109; C1; 1.  
 DR SMART: SM00228; PDZ; 1.  
 DR SMART: SM00233; PH; 1.  
 DR SMART: SM00325; Rhogef; 1.  
 DR PROSITE: PS00094; C5\_MTASE\_1; UNKNOWN\_1.  
 DR PROSITE: PS00479; DAG\_PE\_BIND\_DOM\_1; 1.  
 DR PROSITE: PS50081; DAG\_PE\_BIND\_DOM\_2; 1.  
 DR PROSITE: PS50106; PDZ; 1.  
 KW Photol-ester binding.  
 SQ SEQUENCE 2559 AA; 280946 MW; 7BC661AA1E6EBA0E CRC64;

Query Match 11.7%; Score 922.5; DB 5; Length 2559;  
 Best Local Similarity 20.8%; Pred. No. 1.7e-48;  
 Matches 479; Conservative 279; Mismatches 654; Indels 895; Gaps 82;

QY 17 QSHLSPIASWLSLSLSDSTPERTSPSHRQPSDSTSETAGLVQRC-----VII 67  
 DB 204 QSDNSNPVLQAGERSLNLTLPSRLDSGHTQESTTPATPTSTPLALPKNFQYLTLV 263  
 QY 68 QNDQHFQFTVSGDRIVLVQSVRPGGAAMKAGVKBEDRIKNGMTWYSSHLVEYVKKL 127  
 DB 264 RQDSNGYGMKAVSGDNPFVEYSVKGPGAAELAGLVAGDMILRVNGHEVRLKHPYVGLIK 323  
 QY 128 SGAYVALTLTGSS---PSPGVY-----SGLDQNPVAGVLRV----- 161  
 DB 324 ASTVEYELAVKRQOKLTRPSSVSVPYTPSTPILSGRDKRTASTITGPQVPDSTIKRREMYKIQ 383  
 QY 162 -----NP----- 163  
 DB 384 TLQKMLEQKLEMLERKSDQNNPSYKLSSEANIRKLREQLHQVGAEDAPTIVKLOAAAGNKN 443  
 QY 164 ---TTP-----PP----- 170  
 DB 444 TALLTPNQIOLHSATHSNQOFLHLLHHNNLHNNYPPQOQOAPASTPAFLSLPRLSLS 503  
 QY 171 -----PP----- 170  
 DB 504 LSLGTRKKNTEKDLTSSPFGLTTFDLOQQRMSHOAESMSQSNHQHTTPTSOQFFPHQ 563  
 QY 171 -----PPLP---PPQ-----HITG 182  
 DB 564 QOHREKFGPTSKGNKFLISRLIEEDVPRPLPRQNPRLNDLKNAGNAPSGSHLVA 623  
 QY 183 P-----KPLQDPEVQKATQILMM 202  
 DB 624 PVSDDLRAATPOLNRSQOOLPRSTDNSPSNAKSKRSKIKTKALSDP---KMSQMFLOM 680  
 QY 203 ----- 202

DB 681 ESASAGAAGSIEVDGPPPLPRPLGMMTEDMRSQCNLAQNPSTAFNPLVSTT 740  
 QY 203 -----LRQEEELQDIL-----PPCGETSQR----- 223  
 DB 741 TAVQNDNLNIAFPISGRPNIVQQLQYQQQQQHOMSGGATGALGOTPYLGNKKHRRVGS 800  
 QY 224 -----TCESRLVSQDQADSLDSTERTSPSISELM-----NRNS 259  
 DB 801 SPDNMHRPRPRRTYTKTSSWEIWEKDGSS--PPGTPPPYLSSSHMTVLEDPENNRGA 859  
 QY 260 VLSDDP--LDSPQ-----TSPTLARAQV--HRRQSGDALLPLNHGIDQSPKPLTI 309  
 DB 860 AAAGPCVFIESHQFTPMACASSPIPLSLSNMHAQSD-----TQKEIIS 906  
 QY 310 GPEEDYD-----PGYPNNSDIIFQDLKLSHPAYLVFLRYILSQADPGPLFLYL 361  
 DB 907 MEDENSDLDPEFLDENGPPNNLIRLEAE-----NVTFALAFIVNISNDPAPLFLYL 960  
 QY 362 CSEVYQOTNPQRSLSGLKIMNIFLEKNAPLR-VKIPEMLQAEIDLRLANNEDP-----RN 416  
 DB 961 ITELKESGSKDMRKWAVEIHSTFLVPAPRLSMYRODESLAREVDVLDLQYKYEILRT 1020  
 QY 417 VLCEQEAVALLEIQIINDYRSKRTLGSLYGEND--LLGLDGPRLREOMAERQOL--- 471  
 DB 1021 VELRSKKRAKDLISEQLREFOQRKTAGLSTYIGPDDKLAETKDKLRE-QIIDKYLAPN 1079  
 QY 472 -----AALGDILSKYEEDRSAPMDFA--VNTFMH----- 499  
 DB 1080 LHALIEDENGSPPEQRYKALCSALSTVYRIENRPPSSLYERNHHVSDKSPKSI 1139  
 QY 500 -----AGIR---LRE-----SRSCIAKTOS 518  
 DB 1140 MGNKRMNVRGHPVLYROYEYTHGNHCOTIIGVSPQGYHCTDCKLNIHROSRYVDES 1199  
 QY 519 APDKQKLPFPFKTKKSSNSKKR-----DALEDKR----- 551  
 DB 1200 CGCP-----LQAKRLANDKISKFMKIRPTSDVIGNERSKODELVELPDDGQ 1253  
 QY 552 -----NPILRYIGKPKSSOSIKPGVNRNIIQHFENSHOYDVEPG----- 592  
 DB 1254 ASIVQPSDRPRDANISIRSNQNTSCNTSGL---NTTDQSSFHGSCANDSINPGAGAC 1310  
 QY 593 TORLST-----GSFPE---DLESOSR-----SEIRLGRS 620  
 DB 1311 NMDLSTVAASTPTSTSGVAAGLSAFAELNALDLYDKERRRRRYOHPKHSAPVSNWS 1370  
 QY 621 GSLKREEMKRSRKAEEN-----VPRPSDYDMAAEEAARLHOSASSASLS----- 668  
 DB 1371 ESKYKRLSKRRNRNRKRTSDPSLSRP--NDEQLDLGSLNATYVGSNSLSAGGTESP 1429  
 QY 669 TRSLENPTTP-----FTPKMGRSIESPNL-----GECTDVILPHLLEDLQOLS-- 713  
 DB 1430 STSMEHFAAPGAAGVQVPPMGLNQNHPHLLIQHQAQYCCQODSFOAGLGAASASNA 1489  
 QY 714 -----DLEPEYVQ--NMQHTYGVQVNAVLQREIDROEVINELEYT 753  
 DB 1490 NSSFWNAGHPILVAPRWTLSEDEDDVNEADWSMVAEASALTLTAEEKROEITINEIYQT 1549  
 QY 754 EASHRLTFLVDLIFYORAKENLMPREELARLF--PNLPETELIINSNCEAKKRLR--EG 811  
 DB 1550 ERNNHYRTLKLDRLFLPLYESGLSQDHLLEFPALLSLREINIGARFGQSIKORRIENH 1609  
 QY 812 PIIRDISPMLARFDPAREELQYVAAFCYSQVSALELIRTKOKRESRFQLEQAEASH 871  
 DB 1610 HVVNTIGIDLMDMFGQSGVVLCEFAAQOIALALTEKRRKMDLOKLLKSSSH 1669  
 QY 872 PQCRLOLRDLIVSEMOULTKYRILLENIKHT-----EGGISEHKKLQARQCEILKF 927  
 DB 1670 KACRLELEKDLPLVLORFYKPLPLEFNLYKTVBLRPENTTEAIAIDPAVESSRIIVE 1729  
 QY 928 VNEAVKOTENRRLLEGYQKRLDATALERASNPILAEFFKSLDTTRKMHIEGPITRISK 987  
 DB 1730 VNOAVRTAEDAKHLQNIQRKDRSSYDK-----BEFKKLDLTQNHLLHDGNTL--IKKN 1781





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: September 18, 2002, 10:37:59 ; Search time 16.83 seconds  
(without alignments)  
3513.057 Million cell updates/sec

Title:  
Perfect score:

Sequence: US-09-695-795-4  
1 MSIRLPHSIDRSASKKQSHL.....EGSDIALDSATDFAVSPGP 1527

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	521	6.6	1693	1 RIP2_MOUSE	P97433 mus musculus
2	509.5	6.5	893	1 GEFH_HUMAN	Q92974 homo sapien
3	462	5.9	596	1 LFC_MOUSE	Q60875 mus musculus
4	428	5.4	424	1 LBC_HUMAN	Q12802 homo sapien
5	318	4.0	1805	1 NEST_RAT	P21863 rattus norv
6	262.5	3.3	1356	1 ROM2_YEAST	P70478 rattus norv
7	253.5	3.2	2842	1 APC_RAT	P48681 homo sapien
8	247.5	3.1	1618	1 NEST_HUMAN	Q9UKW4 homo sapien
9	246.5	3.1	847	1 VAV3_HUMAN	P54100 rattus norv
10	245.5	3.1	843	1 VAV_RAT	Q9UC08 mus musculus
11	244	3.1	847	1 VAV3_MOUSE	P27870 mus musculus
12	240	3.0	845	1 VAV_MOUSE	Q75962 homo sapien
13	238.5	3.0	3038	1 TRIO_HUMAN	Q07139 mus musculus
14	237	3.0	738	1 ECT2_MOUSE	P15498 homo sapien
15	233.5	3.0	845	1 VAV_HUMAN	Q60992 mus musculus
16	232	2.9	868	1 VAV2_MOUSE	Q9H8V3 homo sapien
17	232	2.9	883	1 ECT2_HUMAN	P52734 mus musculus
18	231	2.9	960	1 FGD1_MOUSE	Q61315 mus musculus
19	229	2.9	2845	1 APC_MOUSE	P52735 mus musculus
20	227	2.9	878	1 VAV2_HUMAN	P52735 mus musculus
21	227	2.9	2774	1 MAPA_RAT	P34926 rattus norv
22	226.5	2.9	2805	1 MAPA_HUMAN	P78599 homo sapien
23	226	2.9	1591	1 TIAM_MOUSE	Q60610 mus musculus
24	223.5	2.8	1155	1 ROM1_YEAST	P53046 saccharomyc
25	223	2.8	2230	1 GOG4_HUMAN	Q13439 homo sapien
26	221.5	2.8	519	1 ARH5_HUMAN	Q12774 homo sapien
27	219	2.8	2044	1 STEF_DROME	P91620 drosophila
28	219	2.8	2064	1 STEF_MOUSE	P91620 drosophila
29	214.5	2.7	1591	1 TIAM_HUMAN	Q13009 homo sapien
30	213.5	2.7	9614	1 FGD1_HUMAN	Q09763 schizosach
31	211.5	2.7	753	1 YAF9_SCHPO	P30622 homo sapien
32	211.5	2.7	1427	1 REST_HUMAN	Q02952 homo sapien
33	211	2.7	1781	1 AKAC_HUMAN	Q02952 homo sapien

34	209.5	2.7	1174	1 ZO2_CANFA	Q95168 canis famill
35	209.5	2.7	2738	1 PGCV_RAT	Q9EBY4 rattus norv
36	206	2.6	1190	1 ZO2_HUMAN	Q9UGY2 homo sapien
37	204.5	2.6	854	1 CC24_YEAST	P14433 saccharomyc
38	204	2.6	2464	1 MAPA_MOUSE	P14873 mus musculus
39	201.5	2.6	3321	1 KEND_HUMAN	Q95613 homo sapien
40	201	2.5	4687	1 PLEI_RAT	P30427 rattus norv
41	199.5	2.5	2349	1 TPR_HUMAN	P12270 homo sapien
42	199	2.5	1505	1 CURT_HUMAN	P39880 homo sapien
43	197	2.5	2017	1 MYSN_DROME	Q99323 drosophila
44	197	2.5	4473	1 PLEI_CRIGR	Q9J155 citreolus
45	195	2.5	1395	1 SP41_YEAST	P38904 saccharomyc

## ALIGNMENTS

```

RESULT 1
ID RIP2_MOUSE STANDARD; PRT; 1693 AA.
AC P97433;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Rho-interacting protein 2 (Rho-guanine nucleotide exchange factor)
DE (RhoGEF) (RIP2).
GN RHOP2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
ON NCB1_TaxID=10090;
RX MEDLINE=97344280; PubMed=9199174;
RA Gebbink M.F.B.G., Kranenburg O., Poland M., van Horck F.P.G.,
RA Housa B., Moelenar W.H.;
RT "Identification of a novel, putative Rho-specific GDP/GTP exchange
RT factor and a Rho-binding protein: control of neuronal morphology.";
RL J. Cell Biol. 137:1603-1613(1997).
CC -!- FUNCTION: GUANINE NUCLEOTIDE EXCHANGE FACTOR THAT INTERACTS WITH
CC RHOA, BUT NOT WITH RAC OR CDC42. ACTIVATES RHOA TO PROMOTE
CC CYTOSKELETAL CONTRACTION AND INHIBIT NEURITE OUTGROWTH.
CC -!- TISSUE SPECIFICITY: HIGHLY ENRICHED IN THE BRAIN.
CC -!- SIMILARITY: CONTAINS 1 DBL-HOMOLOG DOMAIN (DH).
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.
CC -!- SIMILARITY: TO MAMMALIAN GUANINE NUCLEOTIDE REGULATORY FACTOR LFC
CC AND HUMAN LBC PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
EMBL: U73199; AAB18197.1; -
MGD: MGI:1349443; Rho1p2.
InterPro: IPR002219; DAG_PE-bind.
InterPro: IPR001849; PH.
InterPro: IPR000219; RhoGEF.
Pfam: PF00130; DAG_PE-bind; 1.
Pfam: PF00169; PH; 1.
Pfam: PF00621; RhoGEF; 1.
SMART: SM00109; CI; 1.
SMART: SM00233; PH; 1.
SMART: SM00325; RhoGEF; 1.
PROSITE: PS00479; DAG_PE_BIND_DOM_1; 1.
PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
PROSITE: PS00741; DH_1; FALSE_NEG.

```

DR PROSITE: PSS0010; DH\_2; 1.  
 DR PROSITE: PSS0003; PH\_DOMAIN; 1.  
 KW Guanine-nucleotide releasing factor; Coiled coil.  
 FT DOMAIN 99 266 LEU-RICH.  
 FT DOMAIN 266 310 COILED COIL (POTENTIAL).  
 FT DOMAIN 279 318 POLY-GLU  
 FT DOMAIN 632 698 PHORBOL-ESTER AND DAG BINDING.  
 FT DOMAIN 846 1041 DH.  
 FT DOMAIN 1095 1184 PH.  
 FT DOMAIN 1421 1522 COILED COIL (POTENTIAL).  
 SEQUENCE 1693 AA; 190325 MW; 4E96087C449FF14C CRC64;

Query Match 6.6%; Score 521; DB 1; Length 1693;

Best Local Similarity 20.2%; Pred. No. 9.7e-17;  
 Matches 290; Conservative 218; Mismatches 496; Indels 430; Gaps 55;

QY 205 QEEELQDILPFGERTSQTCERLSDVDSQADSGDSTERTSPSSSELMRNNSVLSDP 264  
 DB 283 EDEAKTEKATPSG-----AAETEEVYRNLESG--RSPSEEE--DAKSIKQ- 326  
 QY 265 GLDSPQTSPIVILARVAQHRRQGSADALLPLNHQIDQS--PKPLIIGPEEDY----DP 317  
 DB 327 -VDGP-----SEH---EDQRLALDRSPDGLKKKHVPASLAAGQLSVLNGGB 372  
 QY 318 GFYN-----NESDIIEDLEKLSHPAYLVVFLRYLSQADPGFLFLYLCSEVYQQTN 370  
 DB 373 VVANCWVIDQVGLDINDYINLEGLSTH-----TS 401  
 QY 371 PMDSRLGKDINWIFLEKNAPLKVKIPMLQAEIDLRLNNEPRN-----VICGAE 423  
 DB 402 PESGRM-----LGPOACHTLPPTDTPSGRPL--IENSEGLDAASQSSTVPTSSSR 452  
 QY 424 AVMLEIQ-----EQINDYRSKRTTLGLSLYGENDLLGLDGPPL--RRROMAKOLAL 474  
 DB 453 TSLNLSFGLHGEKQSHLKKRSSSLDALVADSEGGSEPPICVAVGSSQSPRTGLPS 512  
 QY 475 GDLTSYEEDRSAPMDFAVNTFMS-----H-----AGIRLRSRK--SCTAETQAPDK 522  
 DB 513 GDLDSFEFTNTEPCDNISRTESLSLSTLHKSLSLGSIR--SRYSQSSPKRISGSR 569  
 QY 523 -----DKMLPFPKTKKOSNSK----- 540  
 DB 570 LVADFTVCSTSEORYSQOEPPGKEKRIQEEEMDEVIRAKSSEKXKYSRTSFILMNM 629  
 QY 541 -----KEKDALEDKRRN-----PILRYIGPKS--SSQSITPGNVNIIQH 579  
 DB 630 TSPRNKSKMKNKDKTEKEMNRHQFVPGTFSGVLQCSGCDKTLGKESLQACMK----- 684  
 QY 580 FENSHQ--YDVERPGTGRIS--TGSFPDDLSEDSRSREIRIGRS-----GSLKG 625  
 DB 685 -ANTHGGCKDAVPCTCKKQEKYKNKKPQSLIGSSSVRVVPAPGLSLHPSSMPTGLPAG 743  
 QY 626 REE-----MKRS--RAEANYPRPSDVMDMAAEARLHOSA--SSSASSISTRSLE 673  
 DB 744 RKEFAQVHPLSSVPGTTLSEFRRAVTSLESGDSWRMRSHDELFOGMSGSPSTESF- 802  
 QY 674 NPTPTPRKMGRRSIESPMIGCTVDILPHLLEDDIGQLSDLEPEREVONMOTYGVKDVV 733  
 DB 803 -----MMEDVVDSSLMWD-----LSSDAQEEAEASSLVVDPSFC 837  
 QY 734 ANLTQREIDROEYINELFTVTEASHLTFLVLDLIFYQRRKELMLPRELARFLPLPL 793  
 DB 838 SHOEKVIRQDVIETFLMOTEVNHIQTLIMSEVFRKGMKEELQDLHSTVDKIFPLDEL 897  
 QY 794 IETINSMCEAMKRLREGE-----IIRDISPMPLARDPGARELDQAOVAFCSYQSV 846  
 DB 898 LETHRFFFSMKERQESGAGSRNFVINGIGLIVQGSSEKASMKRITLIGFCFHNE 957  
 QY 847 ALELITRKQKRSRQLFMQAEASHPOCRRLQIRLIVSEMOFLTKYPLLENTIKHTEG 906  
 DB 958 AMSLFLKELQONK-KFONFTIKRNSLLARRRGIPCECILVTQGITKYPYLVERILQYKE 1016

QY 907 GTSEHEKLCRARDQCEIILKFVNEAVKQTEENRRHLEGYCKRLDATALERASNPAAEFKS 966  
 DB 1017 RTEEHRDLKALGLIKMDMAADVLDKSEYKNOKWLEILNKENTYTKLKNHGFRRKA 1076  
 QY 967 LDLTTRKMIHEGFLTWIRISKDTLDQVLLDLVDVLLQROEBRLLLCKHSTAVGSSDS 1026  
 DB 1077 LLSQERALLHDGLVTKYTGRRKDLALLDLVDLFLQEKDKYTF-----AAVQ 1128  
 QY 1027 KOTFSPKLINAVLIRSVATDRAPFIICTSELGPQIYELVALTSSDKNIMMELLEAV 1086  
 DB 1129 KPS---VISLQGLIAREVANEEKGMFLIASAG--PEWEIHTNKSKEENMMRRLOQAV 1184  
 QY 1087 QNATKHGAAPLPIHPSPPGSPRAYOGSTSSRVEINDESVYHTEKEPKKLDGEGPEGR 1146  
 DB 1185 ESCPEEEG-----GRTSE---SDEERKKAERAVAKI-----QQ 1214  
 QY 1147 VODKOLIAQGEPEVQDEEDELRTLPRAPSLDGENGRTROPVLLALGPLMEGLADA 1206  
 DB 1215 COE-IISNODQOICYLEKLN-----IYAEIGEL-----S 1244  
 QY 1207 ALEDVENLRLILWLSLPGHTVKTQAGPEDDLPTPSVVSITSHPMWPGSGAPRI- 1265  
 DB 1245 GFEDVHLPEHLLI-----KP-----DPGEPQOASLL 1271  
 QY 1266 -----SDSTRLA-----RPEGK-----QPEGEDVAVSSLAHLPPR--TRSS 1299  
 DB 1272 AALAREAESLQVAVKASKMGDVSSSESPGGTVLMDPTSDVPAASPASLVTEGTEGR 1331  
 QY 1300 GWDSPSELDRNPAAEASSTEPASVYKVRKVSLLPGGVGAAKVAGSNAIPDSGQ----- 1354  
 DB 1332 GCMVDV-----DGLQGVVTDLAVS-----DAGEVEYR 1359  
 QY 1355 ---SESELSEVEGAQATGNCFFYVSMAPGLDSS-----TEPTGTPSPSQCHSLPA 1403  
 DB 1360 SFGSSGQSEIIQAIQNLITRLVLSLQALTIQDSHTEIHKLVLQORESLAPS---HSFRG 1415  
 QY 1404 WPEEPQYRGVAGGQCSLVRRDVIDIEHTIEQLTIKRLKMDLMEAHRELKLS 1457  
 DB 1416 GPLQODEKSRVLEKQREELA--NIHKLOHOFQOEBQRMRHRTCDQOOREDAQES 1467

RESULT 2  
 GEFH\_HUMAN  
 AC Q92974; Q15079; STANDARD; PRT; 893 AA.  
 DT 30-MAY-2000 (Rel. 39; Created)  
 DT 30-MAY-2000 (Rel. 39; Last sequence update)  
 DT 01-MAR-2002 (Rel. 41; Last annotation update)  
 DE GEF-H1 protein (proliferating cell nuclear antigen P40).  
 GN LFP40 OR KIAA0651.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99074271; PubMed=9857026;  
 RA Ren Y., Li R., Zheng Y., Busch H.,  
 RT "Cloning and characterization of GEF-H1, a microtubule-associated  
 RT guanine nucleotide exchange factor for Rac and Rho GTPases.";  
 RL J. Biol. Chem. 273:34954-34960(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Brain;  
 RC MEDLINE=98403880; PubMed=9734811;  
 RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. X.  
 RT The complete sequences of 100 new cDNA clones from brain which can  
 RT code for large proteins in vitro.";  
 RL DNA Res. 5:169-176(1998).  
 RN [3]  
 RP REVISIONS.

RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohara O.;  
 RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.  
 RN [4]  
 RP SEQUENCE OF 630-893 FROM N.A.  
 RX MEDLINE-89168219; PubMed-246560;  
 RA Reddy A.B., Chatterjee A., Rothblum L.I., Black A., Busch H.;  
 RT Isolation and characterization of complementary DNA to proliferating  
 RT cell nuclear antigen P40.";  
 RL Cancer Res. 49:11763-11767(1989)  
 CC -1- FUNCTION: STIMULATES GUANINE NUCLEOTIDE EXCHANGE OF RAC AND RHO  
 CC BUT IS INACTIVE TOWARD CDC42, TC10, OR RAS. BINDS TO RAC AND RHO  
 CC PROTEINS IN BOTH THE GDP AND GUANOSINE 5'-3'-O-(THIO)TRIPHOSPHATE-  
 CC BOUND STATES WITHOUT DETECTABLE AFFINITY FOR CDC42 OR RAS. MAY  
 CC HAVE A DIRECT ROLE IN ACTIVATION OF RAC AND/OR RHO AND IN BRINGING  
 CC THE ACTIVATED GTPASE TO SPECIFIC TARGET SITES SUCH AS  
 CC MICROTUBULES.  
 CC -1- SUBCELLULAR LOCATION: COLOCALIZES WITH MICROTUBULES THROUGH THE  
 CC CARBOXYL-TERMINAL COILED-COIL DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
 CC BINDING DOMAIN.  
 CC -1- SIMILARITY: COULD BE THE ORTHOLOG OF MOUSE LYMPHOID BLAST CRISIS-  
 CC LIKE 1 (LFC ONCOGENE).  
 CC -1- CAUTION: THE SEQUENCE SHOWN HERE COMES FROM THE FIGURE OF REF.1,  
 CC APPARENTLY THE SUBMITTED SEQUENCE IS WRONG.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC  
 DR EMBL: U72206; AAC97383.1; ALT\_SEQ.  
 DR EMBL: AB014551; BAA31626.2; -.  
 DR EMBL: X15610; CAA33634.1; -.  
 DR InterPro: IPR002219; DAG\_PE-bind.  
 DR InterPro: IPR001849; PH.  
 DR InterPro: IPR000219; RhogEF.  
 DR Pfam: PF00130; DAG\_PE-bind; 1.  
 DR Pfam: PF00169; PH; 1.  
 DR Pfam: PF00621; RhogEF; 1.  
 DR SMART: SM00109; C1; 1.  
 DR SMART: SM00233; PH; 1.  
 DR SMART: SM00325; RhogEF; 1.  
 DR PROSITE: PS00479; DAG\_PE\_BIND\_DOM\_1; 1.  
 DR PROSITE: PS00081; DAG\_PE\_BIND\_DOM\_2; 1.  
 DR PROSITE: PS00741; DH\_1; FALSE\_NEG.  
 DR PROSITE: PS50010; DH\_2; 1.  
 DR PROSITE: PS50003; PH\_DOMAIN; 1.  
 KW Guanine-nucleotide releasing factor; Phorbol-ester binding; Zinc;  
 KW Coiled coil.  
 FT DOMAIN 40 86 PHORBOL-ESTER AND DAG BINDING.  
 FT DOMAIN 234 431 DH.  
 FT DOMAIN 485 571 PH.  
 FT DOMAIN 589 610 COILED COIL (POTENTIAL).  
 FT DOMAIN 797 868 COILED COIL (POTENTIAL).  
 FT CONFLICT 1 21 MSRIESTLRARIDRSRELSK -> LGSAAAGCCCCGCCP  
 FT CONFLICT 193 193 E -> EA (IN REF. 2).  
 FT CONFLICT 579 579 S -> P (IN REF. 2).  
 FT CONFLICT 867 867 Q -> P (IN REF. 2).  
 FT CONFLICT 885 893 SSPQAMPCT -> SLPGAGALYLSFNPPOPSRGTDRLLDPV  
 FT (IN REF. 2).  
 SQ SEQUENCE 893 AA: 101173 MW: 80AC8FA7F762E9C3 CRC64;

Query Match 6.5%; Score 509.5; DB 1; Length 893;  
 Best Local Similarity 23.5%; Pred. No. 1,4e-16;  
 Matches 193; Conservative 148; Mismatches 316; Indels 165; Gaps 29;

OY 594 ORLTGSPFEDLLESDSSRSREIRLGRSGSLKGREEMK--RSRAEYVPRPSVDMDAA 651  
 Db 118 ERPSAIIY-----SDSFROSL-----LGRGRGSSLSLAKSVSTYIAGHND-----E 162  
 OY 652 EAARLHQASSSASSSLSTRSLENPPPTPKMGARSIESPNIGCFCDVILLPHLEDDLQ 711  
 Db 163 SPGLRRILISDSTINMR-----NRT-----LSVES-----LIDEVIYSEL-- 200  
 OY 712 LSLPEEPE-----VQNMQHYGKVVANLQRELDROEVINEFVTASHLRTLRVLDLF 768  
 Db 201 MSDFENDEDFADSWSLVNDSSFLQOHKKEWKQODVYIELQTLHHVTRLTKMTRLF 260  
 OY 769 YQMRKENLMPRELRLEPNLPLELIEHNSWCEAKKIREEG--P-----IIRISDP 820  
 Db 261 RIGMLEELHLEPVGVLPCVDELDIRFLSQLERRRQALCGSTRNRYTHLGLD 320  
 OY 821 MLARFDGPAREELQVAAQFCYSQVSALELITRKORKEHPFQEMQASHPCRRLOLR 880  
 Db 321 LLSQFGSPSAEQCKTYSFECSSRSKALKYKELVARDKRFQOFIRKVTIRPAVLKHHGY 380  
 OY 881 DLVSEMRQRTKYPLLLENIKTKTEGTSHEKLCAROCRRILTFVNEAVKOTENRR 940  
 Db 381 ECTLVTRQRTKYPILTSRILOHSGIEERODLTALGLVKELSNVDEGIYQLEKGR 440  
 OY 941 LEGYQRRLDATALERASNPILAE--FKSLDLTRKMIHESPLTWRTSKDTLDIQLVLE 998  
 Db 441 LGQTYNRMD-----PRAQTPVPGKPGFGRELLARKLIHQCCLMKATNGFKVLLWMT 496  
 OY 999 DLVYLQROBERLLKCHSTAVGSSDKOTSPYKLNANVLRSAVTDKRAFFITCTSE 1058  
 Db 497 DVLVPLQEKDQKYPFLTKPS-----VVSLONLIVRDJANDEKGFLLSAA- 543  
 OY 1059 LGPQIYELVALVSSKNITMELLEPAVONATKHPGAAPRPIHPSPGSOEPAYOGSTSS 1118  
 Db 544 --PPEMYEVHTASRDSTRVIRYIQOSVRCPSREPSLIE-----TEDEVIYLRIM 594  
 OY 1119 RVEINDSEVYHTEKEPKKLEPGGPORVODKOLIAQGEVPO--EEDDEELRLTPRAPPSL 1177  
 Db 595 ELQQRALVELLRE-----KVLGFAEMTHFQAEEDGGSGMALPTLP--- 636  
 OY 1178 DGENRGIRTRDPVLLALGTPLMEGLADALEVENLRHLILMSLPGHTVQQAAGEPE 1237  
 Db 637 ---RGL-FRSESLSPRG-----ERLLQDAIREVEIKLDLY-----GPG 672  
 OY 1238 DDLTPPSVYSIHSHPWDC--SP-----GQAPTISDSTRLARPEGSGEGEDVANVSSLA 1290  
 Db 673 VELLTLPREPALPLEPDGGNTSPVTANGAEATFNGSIELCRADS----- 718  
 OY 1291 HLPPTRRSGVWDPSELDRNPAAEAASTEPAASYKVKVKSLLPG--GGVGAAKVAGSNA 1348  
 Db 719 -----DSSQDRNR-GMQLRSPQELAQRLVNLXGLHGIQAANAQDITLMEAR 765  
 OY 1349 IPDSGQSESELSEFV---EGCAQATGNCFYVSMAPGLDSSTE 1387  
 Db 766 FPEGPERREKLCRANGRDGEAGRAG-----AAVPAPEKQATE 802  
 RESULT 3  
 LFC\_MOUSE STANDARD; PRT; 596 AA.  
 ID LFC\_MOUSE  
 AC 060875; 009115;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE Lymphoid Blast crisis-like 1 (LBC'S first cousin) (Oncogene LFC)  
 DE (RHOBIN).  
 GN LBCL1 OR LFC.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 ON NCBI\_TaxID=10090;  
 RN [1]

```

RP SEQUENCE FROM N.A.
RC TISSUE=Myeloid;
RX MEDLINE=95355462; PubMed=7629163;
RA Whitehead I., Kirk H., Tognon C., Trigo-gonzalez G., Kay R.;
RT "Expression cloning of lfc, a novel oncogene with structural
RT similarities to guanine nucleotide exchange factors and to the
RT regulatory region of protein kinase C.";
RL J. Biol. Chem. 270:18388-18395(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA Olofsson B.;
RL Submitted (JAN1997) to the EMBL/Genbank/DBSI databases.
CC -1- TISSUE SPECIFICITY: UBIQUITOUS, WITH THE EXCEPTION OF LIVER
CC TISSUE. LEVELS ARE HIGH IN HEMOPOIETIC TISSUES (THYMUS, SPLEEN,
CC BONE MARROW) AS WELL AS IN KIDNEY AND LUNG.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAIN.
CC -1- SIMILARITY: TO HUMAN NUCLEOTIDE EXCHANGE PROTEIN LBC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U28495; AAC5234.1; -
DR EMBL: X95761; CA65067.1; -
DR MGI: 103264; Lbc11.
DR InterPro: IPR002219; DAG_PE-bind.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhogEF.
DR Pfam: PF00130; DAG_PE-bind. 1.
DR Pfam: PF00169; PH. 1.
DR Pfam: PF00621; RhogEF. 1.
DR SMART: SM00109; C1; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhogEF. 1.
DR PROSITE: PS00479; DAG_PE_BIND_DOM.1; 1.
DR PROSITE: PS00081; DAG_PE_BIND_DOM.2; 1.
DR PROSITE: PS00741; DH.1; FALSE_NEG.
DR PROSITE: PS00010; DH.2; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
KW Guanine-nucleotide releasing factor; Proto-oncogene;
KW Phorbol-ester binding; Zinc.
FT DOMAIN 40 86 PHORBOL-ESTER AND DAG BINDING.
FT DOMAIN 236 433 DH.
FT DOMAIN 487 573 PH.
FT DOMAIN 493 496 POLY-LEU.
FT CONFLICT 1 21 MSRIESTLRARIDRSKEQATK -> MSGNRROPSPRRGQ
FT CONFLICT 1 21 (IN REF. 2).
FT CONFLICT 156 156 A->V (IN REF. 2).
FT CONFLICT 574 596 MISSING (IN REF. 1).
FT SEQUENCE 596 AA; 68564 MW; E168508BDC7C6E13 CRC64;
SQ

```

Query Match 5.9%; Score 462; DB 1; Length 596;  
 Best Local Similarity 26.2%; Pred. No. 1.3e-14;  
 Matches 119; Conservative 113; Mismatches 173; Indels 50; Gaps 11;

```

DB 257 IMTRIFRTGMLDELQMEPEVVOGLPCVDELSDIHRFLNQLERRRQALCGSTRNFVI 316
QY 815 RDSPMLARFGPAREELQOVAOCYSQVALELIRFKOKRESFOFMQEAESHPOC 874
DB 317 HRLGDLISQFSGSNADQKRTYSEPCSNHTKALKIKLYARDKRFQOFIKMTRSAVL 376
QY 875 RRLQRLDLIVSEMRQTKYPLLENIKHTEGTSEHEKLCRAPDOCRILKFNVEAVQ 934
DB 377 KRHGOECILVTQRTTKYPLINRLIQNSHGVEEYQDLASALGLVKELLSNVDQVHE 436
QY 935 TENRRHLEGYQRIDATALERASNPAAE--FKSIDITRKMIHGPIWRTSKNKTIDL 992
DB 437 LKEARLQETVRMD---PRAQTPYGGPGRGDELRLRKILHRCGLMKTAGTFKDV 492
QY 993 QVLLLEDVLLQROBERLLKCSHTAVGSSDSKQTFSPVLKNAVLRVATDKRAEF 1052
DB 493 LLLMTDVLVFLQERDQKIFPSLDKPS-----VYSIANLYRIQIANQAKMF 540
QY 1053 IICTSELGPPQIYELVALTSSDKNIMLEEAQV 1087
DB 541 LISS---GPPEMYEVMAASRDDRTWIRVIOQSVR 572

RESULT 4
LBC_HUMAN STANDARD; PRT; 424 AA.
AC 012802;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE LBC oncogene (P47) (Lymphoid blast crisis oncogene).
GN LBC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94119604; PubMed=8290273;
RA TOKSOZ D., Williams D.A.;
RT "Novel human oncogene lbc detected by transfection with distinct
RT homology regions to signal transduction products.";
RL Oncogene 9:621-628(1994).
CC -1- FUNCTION: STIMULATES EXCHANGE ACTIVITY ON RHO PROTEINS IN VITRO,
CC BUT NOT ON CDC42, RAS OR RAC. MAY BIND CALCIUM IONS.
CC -1- TISSUE SPECIFICITY: RESTRICTED TO HEMATOPOIETIC CELLS AND SKELETAL
CC MUSCLE, LUNG AND HEART. NOT FOUND IN BRAIN, PLACENTA, LIVER,
CC PANCREAS OR KIDNEY.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
CC -1- SIMILARITY: TO MAMMALIAN LFC ONCOGENE AND TO MOUSE RHO-GUANINE
CC NUCLEOTIDE EXCHANGE FACTOR (REF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U03634; AAC50065.1; -
DR MIM: 605396; -
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhogEF.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhogEF. 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00325; RhogEF. 1.
DR PROSITE: PS00741; DH.1; FALSE_NEG.

```



```

QY 766 LIFYQMRKKNIMPREELARLPNPLLEIHNHSMCEANKKLEEGPIIRIDISDPMIARR 825
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 818 V-----EEEDQRIVKPLEKVSQ-----DSLGLAEENV----- 845
QY 826 DGPARELQOVAQFCYSVALERIKOR-----ESR---FQFMQAEHPQCRKL 877
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 846 -----QPLRYLEEDDCINS-----LEDKTHKSLGLEDRNGDSIIIPDESETOVSLRPP 896
QY 878 QLRDLIVSEMRUTKYPPLLENIKTEGTSHEKLCARADOCREILKFVNEAVKOTEN 937
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 897 FEED-----QRIVNLEKESQEFSSSEEBE-----VMERSLEGEN 933
QY 938 RHRLEGYQ--RIDATALERASNPLAEFKSLDTTRKMIHEGPLTWIRISKDXTLDQVL 995
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 934 HESLSVEKEQVMSQLEKESQDSG---KSLEDESEQETF--GPL-----EKEVAESLRSL 984
QY 996 LLEDVLVLLQROEERLLCHKSTANVGSSDSKOTFSPVLKINAVLRSVATDKRAFIIC 1055
DB 985 AGOD-----OEOKLEDETOOTLRAVQEQMAVSPREKVDLEKPLGNDQ-----I 1032
QY 1056 TSELGPPQIYELVALTSSDKNIV-----MELLEAVONATKHPGAPI----- 1098
DB 1033 ANSLGKENGESIVSL--KEKGIEYKSLFETIETLEFABEEDLERKKSIDTOEPLMSTEV 1090
QY 1099 -----PIRPSPPGQEPAYQGSTSRVLEINSEVYHTEKEPKLPE--GP--GPEQVQOK 1150
DB 1091 ARETYEPPEDEPPGS-----LGS---VDENRETLTSLKEQSELSSLKWVNETVEQDS 1141
QY 1151 QLIAGEP--VOEDEEELRTLPRAPPSLDGENR-----GIRTRDP 1189
DB 1142 QOCLOYEEGLQROHESLREYKQELPSSGNQORMEDVYBKAQVGEAPLATTVGTEOK 1201
QY 1190 VLLATLGPLMELGLADALEVDENLRHL--LMSLLPGHTVKTQAAGEPEDLTP----- 1242
DB 1202 AELHLRG-----OGGEERAAEGELLQDIVGEANSL-----GSSEPKQGVPAEALD 1248
QY 1243 -----TPSVYTSIHPWDGSPGOAPITISDTRL--ARPESQGEQEVANVSSLA 1290
DB 1249 NIEGALVPAVQAQMEVEREDRAQAQEDQSIETVLGLEAARGLELEQEVGLEDRP 1308
QY 1291 H-----LPPRTSSGVMSPELDNRPAEAASSTEPASVYKVRKSLPGGVAKA 1344
DB 1309 HFAREARIP-----SLGE--ESYAKAIAQGLEPKER-----KEGALDSGLLELPKIS 1357
QY 1345 GSNALPDSQSESESEVSEVGAQATGCFYVMPAGPLDSSTEPRTGPPSPSQ----- 1397
DB 1358 -SEALECQGHESSE--SMEGWEEB-----ASLETSDHGSDAPQPRPETEDECQAQA 1409
QY 1398 -----CHSLP-----AMPTERQPYRGVR-----GGQCSLYRRDQVDFHTI 1434
DB 1410 LTAPEGKLLPECPPIITLDAHELOQA--EGIQAGWOPDASEALERVENEPFGUGEI 1468
QY 1435 EOLTIKHLKLMELAHRELKLSLGESSGTTPVGSPFHTAA--RWT---DYSLSP--- 1486
DB 1469 PE-----GLQWMEBERESEADLDGELTLPDSTPLGLYLRSPASPKWDLADGQRSLSPQD 1522
QY 1487 -----PAKEALASDSQNGQEGQSCPEEGSDIA---LEDSATDTVAVSPG 1526
DB 1523 AKGEDMGAVPAPAAQGLSGPREEEEOG---HGSDLSSSEFEDLTGEASILPG 1571

```

```

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972.
RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
RA Favello A., Fulton L., Gattung S., Greco T., Kirsten J.,
RA Kucaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
RA Marlis E., Menezes S., Miller N., Nhan N., Pauley A., Peluso D.,
RA Rifken L., Riles L., Tatch A., Trevisan E., Vignati D.,
RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.,
RL Submitted (JAN-1995) to the EMBL/Genbank/DBCP databases.
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=96208506; PubMed=8641285;
RA Ozaki K., Tanaka K., Imamura H., Hihara T., Kameyama T.,
RA Nonaka H., Hirano H., Matsura Y., Takai Y.;
RT "Romlp and Rom2p are GDP/GTP exchange proteins (GEPs) for the Rho1p
RT small GTP binding protein in Saccharomyces cerevisiae."
RL EMBO J. 15:2196-2207(1996).
CC -I- FUNCTION: STIMULATES THE EXCHANGE OF RHO1 GDP-BOUND FORM INTO
CC GTP-BOUND FORM.
CC -I- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: 019103; AAB67564.1; -.
DR SGD: S0004363; ROM2.
DR InterPro: IPR001180; CNH.
DR InterPro: IPR000591; DEP.
DR InterPro: IPR000219; RhGEF.
DR Pfam: PF00780; CNH; 1.
DR Pfam: PF00610; DEP; 1.
DR Pfam: PF00621; RhGEF; 1.
DR SMART: SM00036; CNH; 1.
DR SMART: SM00049; DEP; 1.
DR SMART: SM00325; RhGEF; 1.
DR PROSITE: PS00741; DH_1; FALSE_NEG.
DR PROSITE: PS50010; DH_2; 1.
DR Guanine-nucleotide releasing factor.
FT DOMAIN 659 846 DH.
FT DOMAIN 252 265 POLY-ASN.
FT DOMAIN 329 336 POLY-HIS.
FT DOMAIN 632 635 POLY-ASP.
SQ SEQUENCE 1356 AA; 152595 MW; 5FBC54211AE7BC92 CRC64;

```

Query Match 3.3%; Score 262.5; DB 1; Length 1356;  
Best Local Similarity 20.1%; Pred. No. 6.7e-05;  
Matches 200; Conservative 151; Mismatches 311; Indels 331; Gaps 46;

```

QY 235 EADSGIDSGTERFPFSISESLMNRNSVLSDPGLSDSPQTSFVILIARYAONHRRGSDAALLP 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 33 DSDSHGDISQLLPRRIENIQNLNVLLSE---DIANDIT---IAKORRRGVEAA--- 81
QY 295 LNHGIDGSPKPLIIGPEEDYDGYFNNE-----SDIIFQDEIKLSHPAYLVFLRYIL 349
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 82 -----IDSDIP-----NNEMKGSNYILISQOTNIKEVPD-----TQSL 116
QY 350 SGADGPGPLFLYLCSEVYQOTNPKDSRL-----GQDMMIFLEKAPLRY----- 394
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 117 SSADMTPV-----SSPKRADATSSHPIVHAKSMSHIYSTSNASQAQAHYNDH 165
QY 395 KIPKEMIQAEIDLRLNNEDPRNVLCFAOEAVML----- 427
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 166 PLPPM-----SPRNEYQKNKSTYAFVPKRKRSLPOLALAGLKKQSSFSFGS 212

```



```

Oy 428 -----EIOEO-----INDYRSKRTIGLSGLY--- 448
Db 213 ASTTQARKSPLOGFFGFFSPRSSKDLHGOHNOHIOHNNINNNNNNTNNGAHYOVG 272
Oy 449 GENDLLGDGDPLERRORAEKOLAAGDILSKYEEDRS-----APMDFAVNTFM----- 497
Db 273 SSNSNYPOHSHSISRSMSLNS-STLKNIASSFQSKTSNSRKATQKYDITSNPFSDPHH 331
Oy 498 -----SHA-----GIRLRESRSCSTAK-----TOSA 519
Db 332 HHHHSSSHSLNNVHSGNSSSVMGSSSSNIGLTKTRVASTSLALRKRTSVSGSTLSS 391
Oy 520 PDKMKWLPFFPKTKKSSNKKER--DALEDKRRNPILRTYIGKPPSSSO--SIRPGNV- 573
Db 392 PRSSMTPLASRPVMASSKKPOVYPAL-----LSRVATKFKSSIQLEGHKKDGLVY 444
Oy 574 -----RNIOHPENSHQ-----YDVEPQTORLSTGSEPEDLE 607
Db 445 RDAFTGQAVDVICALIRTSDBRNLLAFGRSLDAQKLFHDVYEHRLDS---PHEVYE 500
Oy 608 -SDSRSTIRLRGSGSLKGREM-----KRSKKAENVPRPSDV-----DMDAA 651
Db 501 FTDMS---RFTGTGTNAHDPMLLPNSSSFGNSHSYPNMSGVSSSTSLNSDQATL 556
Oy 652 EAARLHOSASSASSLSTRLENPTPPT-----PKMGR-----RSIE----- 689
Db 557 TGSRLHSSSLSQOK-NPAALHNHNGVTTLAECSPTCTDALCYSTSCRRLEQOARL 615
Oy 690 --SPNLCFTDVLPHLEDDLGOLSDLEPEPEVONMOHTGVKDVANLTOREIDROEVI 747
Db 616 NLKPGKGLKRNISM--ALDDD-----DEEKPSMTSSVKEDMENLPKKEIKROAI 664
Oy 748 NELPTEASHLRLVLDLIFYQMRKKNLMP-----REELARPENLDELLEIHNMSCEA 803
Db 665 YEVIITENFNFKSLKETDTPTKTLAETNIIISADIRKPIKHNVEFHNIDIVSVNRFLKA 724
Oy 804 MKKLREBPIRIDISDPMILARFDGPARBELQOVAQFCSY--GSVALELIRTKORKESE 860
Db 725 LTRGRSSPVYVRGIGDIYL-RPI-PFFE-----PFVSYASRAYAYVLETOQSVAPY 775
Oy 861 FOLEMOEAESHPOCRRLDRLIVSEWORLTKYPLLENIITKH--EGGTSEHEKLCAR 918
Db 776 FARFDDDDMS--SSLRHGIDSEFLSGVSRPGRYMLLVEIKMSTPEDEKSDYEDLSKAM 833
Oy 919 DQCPREILKFVNAVQETNRRHLEGOYKRLDATALERASNLAEFKSLDLT--TRKMIH 976
Db 834 DALRDEMKRIDQASGAADRDHVKLKKOTI-----LFKNRYVNLGLNDERRKIKH 883
Oy 977 EGPPLTWR-ISK-DKTL--DLQVLLLEDLVLLIQ 1005
Db 884 EGISLRKELSKSDGTVEGDIQFYLLDNMLLFLK 916

```

```

RT "cDNA cloning of the rat APC gene and assignment to chromosome 18.";
RL Mamm. Genome 6:746-748(1995).
RN [2]
RP MORGANESIS.
RC STRAIN=SPRAGUE-DAWLEY, AND FISCHER 344/N;
RX MEDLINE=95148647; PubMed=7846077;
RA Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K.,
RA Weisburger J. H., Sugimura T., Nagao M.;
RT "Specific 5'-GGA-3'-->5'-GGA-3' mutation of the APC gene in rat colon
tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).
CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-
CC CATEININ. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE
CC ALLOWS THE DOWN-REGULATION OF CYTOSOLIC BETA-CATEININ (BY
CC SIMILARITY).
CC -1- SUBUNIT: FORMS HOMODIGOMERS AND ASSOCIATES WITH CATEININS (BY
CC SIMILARITY).
CC -1- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
CC EMBL; D38629; BAA07609.1; -.
CC HSSP; Q02248; 3BC7.
CC InterPro; IPR000225; Armadillo.
CC Pfam; PF00514; Armadillo_seg.6.
CC SMART; SM00185; ARM.5.
CC PROSITE; PS50176; ARM_REPEAT.1.
CC Anti-oncogene; Phosphorylation; Coiled coil; Repeat.
CC KW DOMAIN 1 728
CC FT DOMAIN 1 62
CC FT DOMAIN 125 260
CC FT REPEAT 451 493
CC FT REPEAT 503 545
CC FT REPEAT 546 589
CC FT REPEAT 590 636
CC FT REPEAT 637 681
CC FT REPEAT 682 723
CC FT REPEAT 724 765
CC FT DOMAIN 739 2831
CC FT DOMAIN 1130 1155
CC FT DOMAIN 1356 1575
CC FT DOMAIN 1864 1891
CC FT MUTAGEN 523 523
CC FT SEQUENCE 2842 AA; 310530 MW; 3CB2EAB8A34EBF47 CRC64;

```

Query Match 3.28; Score 253.5; DB 1; Length 2842;  
 Best Local Similarity 19.68; Pred. NO. 0.00047;  
 Matches 358; Conservative 231; Mismatches 717; Indels 517; Gaps 83;

```

Oy 8 SIDRASKKOSHLSPISWLSLSLSDSPERTS---PSHHNRPDPT--SEPTAGLVQ 62
Db 1217 SAPSSNAKRQSQLH-----PSSAQNRNGQTPKAGCAVPSINDTQMTQYCEVEDPPIES 1269
Oy 63 RCVLIIQKDOHGFQFTVSGDRIVLYOVSRPGAA--MKAGVKEGD-----RIIKV 109
Db 1270 RCSSLS-----SLSSADELIGQDTTQDEASANTLQIAELKENDVYRTSAQDPASDVAV 1323
Oy 110 NGMTWYTNSSHLEVYKLI-----KSGAYALTLGSSPPS----- 143
Db 1324 SOSRTKPSRLQASGLASESARHKRAVERFSSGAKSPSKSGAQT-----KSPREHYVDET 1378
Oy 144 -----VGVSGIQ--QNSVAVGLRVNP-----IIP-----PP 173
Db 1379 LVFSKCTSVSLDSFESNSIASVQSEPCGMVSGIVSPSPLDPSPGQTMPPRSKRTIPP 1438

```





DR	PROSITE; PS50003; PH.DOMAIN; 1.
DR	PROSITE; PS50001; SH2; 1.
DR	PROSITE; PS50002; SH3; 2.
KW	Phorbol-ester binding; Zinc; SH2 domain; SH3 domain; Repeat;
FT	Guanine-nucleotide releasing factor; Alternative splicing.
FT	DOMAIN 1 119 CH.
FT	DOMAIN 192 371 DH.
FT	DOMAIN 400 502 PH.
FT	DOMAIN 514 562 PHORBOL-ESTER AND DAG BINDING.
FT	DOMAIN 592 660 SH3 1.
FT	DOMAIN 672 766 SH2.
FT	DOMAIN 788 847 SH3 2.
FT	VARSPLIC 1 107 MEPRQCAQWLHCKVLPITNHRVTMSAQVEDLAQTLDYGV
FT	LICOLLNLRAHSINLKEINRPMOSGLCNKTRTFAC
FT	CETGMRSELFEEAFDELDVDVDFGK -> MQLPPCPRAHL
FT	P (IN ISOFORM BETA).
FT	K -> E (IN REF. 2).
FT	Y -> H (IN REF. 1; AAD20348).
FT	T -> S (IN REF. 2).
FT	V -> A (IN REF. 1; AAD20348).
FT	CONFILCT 429 429 V
SQ	SEQUENCE 847 AA; 97775 MM; CIE29F0B094CB721 CRC64;
<hr/>	
Query Match	3.1%; Score 246.5; DB 1; Length 847;
Best Local Similarity	20.7%; Pred. No. 0.0002;
Matches 142; Conservative 99; Mismatches 224; Indels 211; Gaps 27;	
Dy	701 LPHLEDDLGSLDPEPEYON--MHTYGKVANTFOREIDROE-----YIN 748
Dy	144 LPDLDEL-----VEDEDLYDCYGGDEGEYEEDLMKAEMANRPCPENDIRSCOLA 198
Oy	749 ELFYTEASHLTAKVLDIIFYQRMKENMLPREELARFPNLPPELIETHNSCEAMKKLR 808
Dy	199 EIKTTEEYTTLSIEKYFMAPLKR--FLTAEPDSVFIMPELVKLHRN----- 247
Oy	809 EEGPIIRDSPMLARFGPARRELOV-----AAQCSYSVALLETIRTKORKE 858
Dy	248 ----LMQIHHSIYNKN----QNLYQVFINKEELVIYGYCGSGVEAISISLDYISIKT 299
Oy	859 SRFOLEMOAESHPCCRLQDLDLIVSEMQRTKYPLLENIIKTHEGTSHEKLCRAR 918
Dy	300 EDVALKLEECSKRANNNGFFTLDLVVVMQVKLKHLLLOELVYKTHDPDT--EKANLKAL 358
Oy	919 DQCBEILKFVNEMAKQOTNRHREGYOKRLDATALERASNP-----AAEFKSID 968
Dy	359 DAMKDIAOVNEVKRDNETLEIKOFO-----LISTENLNOPYLLFRGQGGEIRITTLD 413
Oy	969 LTRKMRIHGFLWRISKDKTLDLOVLLLEDVLLLOBOEBRLLLCHSKSTAVGSSDSKO 1028
Dy	414 KHTKOERH-----IFLFDAVICYCRKGDNVEMK----- 442
Oy	1029 TFSYVLKIANVLI-RSVATDKR-----AFFLICSELGPQIYELVALTSSDKRIMMEL 1081
Dy	443 --EIIDQQYKIANNPTTDKENRKWSYGFLLIHQONGGLEFY--CKTYDLKKRWLEQ 496
Oy	1082 LEEAUONTAKHPGAAPIRIPHSPPGSQEPAAVGSTSSVEI----- 1122
Dy	497 FEMALSN-----IRPDYADSNHDHKMHTFTFTVSCYQMILLRGTFVOGYLCF 545
Oy	1123 -----NDSEVYHTEKEPKKLPDEGRPEODRVODKOLIAAGEVQEEDBEL 1167
Dy	546 KCGARAHKECLGRAVNCGRVNSGEGQTLKPEKRTNGLRTPRKQ-VDPGLPKM---QVI 600
Oy	1168 RTLRARPASLDGKERGITRTRPVLLALTGLLMG--LADAALDEVCLR--HLILMSL 1222
Dy	601 RNYSCTRP-----PALHBSPPLQIOAGDVVELLKDAHSLFWQ- 638
Oy	1223 LPGHTVTOAGE-----PEDDLTPTPSV---VSITSHEWPDGSGQAFTISDSTRLARPE 1275
Dy	639 -----GNRIASGVEGFPDSADVRCPCPKRVNVDSIQGWYAGA-----ME 678
Oy	1276 GSQPEGEDV-AVSSLHLPPRTBSG 1300

Db 679 RLQAEELLINRVNSTYLVRHRTKESG 704

RESULT 10

VAV\_RAT STANDARD; PRT; 843 AA.

ID VAV\_RAT

AC P54100;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE VAV proto-oncogene (P95).

GN VAV1 OR VAV.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxId=10116;

RN [1]

RP SEQUENCE FROM N.A.

RA Rivera J., Halem-Smith H.;

RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: PROBABLE EXCHANGE FACTOR FOR A SMALL RAS-LIKE GTP-BINDING PROTEIN. CAN BE ACTIVATED BY TRUNCATION OF THE N-TERMINUS.

CC -1- SIMILARITY: CONTAINS 1 CALPONIN-HOMOLOGY (CH) DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).

CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG BINDING DOMAIN.

CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.

CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.

-----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: U39476; AAA98606.1; -.

DR HSSP: P29354; IGR1.

DR InterPro: IPR003247; CH\_type.

DR InterPro: IPR001715; Calponin\_hom.

DR InterPro: IPR002219; DAG\_pe-bind.

DR InterPro: IPR001331; GDS\_CDC24.

DR InterPro: IPR001849; PH.

DR InterPro: IPR000219; Rhogef.

DR InterPro: IPR000980; SH2.

DR InterPro: IPR001452; SH3.

DR Pfam: PF00307; CH; 1.

DR Pfam: PF00130; DAG\_pe-bind; 1.

DR Pfam: PF00169; PH; 1.

DR Pfam: PF00621; Rhogef; 1.

DR Pfam: PF00017; SH2; 1.

DR Pfam: PF00018; SH3; 2.

DR PRINTS: PR00008; DAGPEDOMAIN.

DR PRINTS: PR00452; SH3DOMAIN.

DR PRODOM: PD001527; CH\_type; 1.

DR SMART: SM00109; C1; 1.

DR SMART: SM00033; CH; 1.

DR SMART: SM00333; PH; 1.

DR SMART: SM00325; Rhogef; 1.

DR SMART: SM00326; SH3; 2.

DR PROSITE: PSS0021; CH; 1.

DR PROSITE: PSS00479; DAG\_pe BIND\_DOM\_1; 1.

DR PROSITE: PSS0081; DAG\_pe BIND\_DOM\_2; 1.

DR PROSITE: PSS0010; DH\_1; 1.

DR PROSITE: PSS0074; DH\_2; 1.

DR PROSITE: PSS0003; PH DOMAIN; 1.

DR PROSITE: PSS0001; SH2; 1.

DR PROSITE: PSS0002; SH3; 2.

DR Proto-oncogene, Phorbol-ester binding; Zinc; SH2 domain; SH3 domain;

KW Guanine-nucleotide releasing factor; Repeat.

FT DOMAIN 1 119 CH.

FT DOMAIN 194 373 DH.

FT DOMAIN 402 504 PH.

FT DOMAIN 516 564 PHORBOL-ESTER AND DAG BINDING.

FT DOMAIN 615 658 SH3 1.

FT DOMAIN 669 763 SH2.

FT DOMAIN 780 840 SH3 2.

SO SEQUENCE 843 AA; 97953 MW; C4A5CAC045FCB80E CRC64;

Query Match 3.18; Score 245.5; DB 1; Length 843;

Best Local Similarity 21.84; Pred. No. 0.00022;

Matches 143; Conservative 90; Mismatches 254; Indels 169; Gaps 26;

QY 700 ILPHLLDD--LG-----OLDLEPEPE-----YQNMQHNGKDVYANLQREI-- 741

Db 126 IMPEPTDSALGDEDIYSGLSDDQIDTAEEDEDLYDCVEN--EEAGEGIEDLRSESVP 184

QY 742 -----DROEVINELFVTEASHLRTLYLDIFYQRMKREMLMPREIARLPNLP 792

Db 185 TPPKMTEYDKKCCCLREIQOTEETKYSIQQHFMKPLQR--FLKPDOMETFVINEE 242

QY 793 LIETHNSMCEAMKRLREBGPRIIDISPMILARDGPAREELQ-----VAAQFCS 842

Db 243 LLSVTHFR--LKLKLD-----ALSGPATMLYGVFIKKERFLVYGRYCS 285

QY 843 YQSVALELIRTKOKESRFOLFQEAESHPOCRLOLRDLIVSEMRITFYPLLENIR 902

Db 286 QVESAIKHLDDVATAREDVQKLEECGQRANNGRFTIRDLMPQMQLYHLLQELVK 345

QY 903 HTEGTSSEHEKLCRAPDQCEILKFVNEAVKOTENRRHLEGYQ---RLDATALERASNP 959

Db 346 HTQ-DTTEKENLRALDAMRDIAQVNEVRKNETLRQITNFQSLINDQSLANVGRPK 404

QY 960 LAEFKSLDITTKMHIEGPITRKISKDKTLDLYLLLEDLYVLQOEERLLKCKSKT 1019

Db 405 IDGELKITSVER--SKT--DRYAFLLDALALKKRGSGSYDLKAS-- 446

QY 1020 AVGSDDSKQTFSPYLKINAVLIRSATDKR-----AFPICTSELPQIYELVLT 1072

Db 447 -----VNLHSFOVRDSSGERDNKKMSMFLI---EEOGAQGIELFPKTR 489

QY 1073 SDKNIMELLEAVONATKHPGAIPRIHSPSGQSPA-----YQG---- 1114

Db 490 ELKKMMEQEFMAISNI--YPENATANGHPQMFSEFTTSCKACQMLLGTFFQGRYCY 547

QY 1115 -----STTSRYEINDSEVYHTEKEPKKLPEGPGEQYVQDKOLTAQGEPEVQEBEE 1166

Db 548 RCRAPAHKECLGRVPPCGRQDFSGTMRKDL-----HRAADKKRNEELIPKREVOGEY 601

QY 1167 LRTLPRAAPSIDGENRGIRTRDPVLLALGTPLMEGLADA--LEDVENLRLHLIWSLT 1223

Db 602 YGIRP--PPGAFG-----PFLRLNPGDIYELTAEBEHMTWEGRNATNEVGH--F 648

QY 1224 PGHTVKTQAAGEPEDDLPPTSVSVTSHPW-----DPSPGQAPPTISDSTRLAR 1273

Db 649 PCNRVRYVHGPPQD-----LSVHLMYAGPMERAGAGILTRNSDGTLYVR 694

RESULT 11

VAV3\_MOUSE

ID VAV3\_MOUSE STANDARD; PRT; 847 AA.

AC O9ROC8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE VAV-3 protein.

GN VAV3.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxId=10090;

RN [1]



[2]  
SEQUENCE OF 1-93 FROM N.A.  
MEDLINE-91172176: Pubmed-2005887;  
RA Ketzav S, Cleveland J L, Heslop H E, Pulido D;  
RT "Loss of the amino-terminal helix-loop-helix domain of the vav proto-  
oncogene activates its transforming potential.";  
RL Mol. Cell. Biol. 11:1912-1920(1991).  
CC -1- FUNCTION: PROBABLE EXCHANGE FACTOR FOR A SMALL RAS-LIKE GTP-  
BINDING PROTEIN. CAN BE ACTIVATED BY TRUNCATION OF THE N-TERMINUS.  
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED IN HEMATOPOIETIC CELLS BUT  
CC NOT IN OTHER CELL TYPES.  
CC -1- SIMILARITY: CONTAINS 1 CALPONIN-HOMOLOGY (CH) DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG  
BINDING DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -1- SIMILARITY: CONTAINS 2 SH3 DOMAINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL: X64361; CAA45713.1; -;  
DR EMBL: M59833; AAA63402.1; -;  
DR PIR: A39576; A39576.  
DR PIR: S36941; S36941.  
DR HSSP: P29354; 1GRI.  
DR TRANSFAC: T01230; -;  
DR MGD: MGI:98923; Vav.  
DR InterPro: IPR003247; CH\_type.  
DR InterPro: IPR001715; Calponin\_hom.  
DR InterPro: IPR002219; DAG\_pe\_bind.  
DR InterPro: IPR001331; GDS\_CDC24.  
DR InterPro: IPR001849; PH.  
DR InterPro: IPR000219; RhoGEF.  
DR InterPro: IPR000980; SH2.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF00307; CH; 1.  
DR Pfam: PF00130; DAG\_pe\_bind; 1.  
DR Pfam: PF00169; PH; 1.  
DR Pfam: PF00621; RhoGEF; 1.  
DR Pfam: PF00017; SH2; 1.  
DR Pfam: PF00018; SH3; 2.  
DR PRINTS: PR00008; DAGPDOMAIN.  
DR PRINTS: PR00452; SH3DOMAIN.  
DR ProDom: PD001527; CH\_type; 1.  
DR SMART: SM00109; C1; 1.  
DR SMART: SM00033; CH; 1.  
DR SMART: SM00233; PH; 1.  
DR SMART: SM00325; RhoGEF; 1.  
DR SMART: SM00252; SH2; 1.  
DR SMART: SM00326; SH3; 2.  
DR PROSITE: PS50021; CH; 1.  
DR PROSITE: PS00479; DAG\_pe\_bind\_dom\_1; 1.  
DR PROSITE: PS50081; DAG\_pe\_bind\_dom\_2; 1.  
DR PROSITE: PS50010; DH\_2; 1.  
DR PROSITE: PS00741; DH\_1; 1.  
DR PROSITE: PS50003; PH\_DOMAIN; 1.  
DR PROSITE: PS50001; SH2; 1.  
DR PROSITE: PS50002; SH3; 2.  
KW Proto-oncogene; Phorbol-ester binding; Zinc; SH2 domain; SH3 domain;  
KW Guanine-nucleotide releasing factor; Repeat.  
FT DOMAIN 1 119 CH.  
FT DOMAIN 194 373 DH.  
FT DOMAIN 402 504 PH.  
FT DOMAIN 516 564 PHORBOL-ESTER AND DAG BINDING.  
FT DOMAIN 617 660 SH3 1.  
FT DOMAIN 671 765 SH2.

FT DOMAIN 782 842 SH3 2.  
FT CONFLICT 29 29 0 -> E (IN REF. 2).  
SQ SEQUENCE 845 AA; 98136 MM; 3666DCDDC1C5229DA CRC64;  
  
Query Match 3.0%; Score 240; DB 1; Length 845;  
Best local similarity 21.4%; Pred. No. 0.0004;  
Matches 168; Conservative 93; Mismatches 300; Indels 224; Gaps 32;  
  
QY 700 ILPHLEDDL-----GQLSDLEPPE-----VQWQITVKGDVYANLTQREI-- 741  
DB 126 IMPPTEDSALNDEDIYSGLSQIDIDTAEDDEDLYDCVEN-EAEGDEIYEDIMRLSESV 184  
QY 742 -----DROEYINLEFTEASHLRTLVLLIFQYRKREMLPREELAFPLNPE 792  
DB 185 TPFRKMTYDKRCCCLRIQQTTEKYTTTLLSIOHFKPLQYR-FLKQDMETIFVNIIE 242  
QY 793 LIEIHNSCEAMKRLREBPRIINDISDPMLARDPAREELQO-----VAAQFCS 842  
DB 243 LFSVHTHF-----LKEKDALA-----GPGATTLVQVFTKYKREFLVYGRYCS 285  
QY 843 -YQSALELIRTKORKSRQLFMQEAESHQCRRLQRLIVSEMQRLTKYLLLENTI 901  
DB 286 QVESASAKHLQVATYARD-VQMKLECSQRANNGRFTLRDLVWPQRYLVKYLHLQELV 344  
QY 902 KHREGGSEHEKLCARDOQREILKFNVEAVKQTENNHRELEGO---KRLDATALERASN 958  
DB 345 KHTQDAT-EKENLRALDAMRDLAQCVEYKRNKNETLRITNQLSTENLDQSLANTGRP 403  
QY 959 PLAAEFKSLDLTTRKMIHEGRLTWIRSKDTLDLVLLLEDLVLLQROBERLLKCHSK 1018  
DB 404 KIDGELKITSVERR-----SKT-DRYAFLLDKALLICKRQSDYDLKAS-- 446  
QY 1019 TAVGSSDSKQTFSPVLKLNVLRSVATDKR-----AFFITCTSELGPQIYELVALT 1071  
DB 447 -----VNLHSFOYRDSGGERDNKKSHMFLI--EDQAGQYELFEFKT 488  
QY 1072 SSDKNIMELLEBAVONATKHPAARPIHPSPGSOEPA-----YQGSTS 1117  
DB 489 RELKKMKMEQFMAISNI--YPENATANGHPQMFSEETSKCAQOMLIRGTFYGYCR 546  
QY 1118 SRVEI-----NDSEVYHTEKEPKRLDEGPGPEGRVODKQIIAAGEPVOEE 1162  
DB 547 YCRAPAHKCEICLRVPCGRHGFAGTKKDKL-----HRRADQKKNELGLPKMEV 599  
QY 1163 DEBELRTLPRAPPSLOENNGITRPDVLLATGLPLMEGLAQA---LEDVENLKHLL 1219  
DB 600 FOEYVGIRP--PGAFG-----PFLRLNPGDIVELTKAEAEHNMWGCRTATNEVG 648  
QY 1220 WSLPLGHVTQAAGEBEDDLTPTPSVSITSHPW-----DQSPGQAPRTISDTRLAR 1273  
DB 649 W--FPCNRVHPYVHGPRQD-----LSVLMTAGPMEERAGAGCILLNRSDDGYTLVR 696  
QY 1274 -----PEGQPEGEDEVAVSSLAHLPPRTSSGCVMS-DE 1306  
DB 697 QRVKDAEFAISIKYVNEVKIKIMTSEGLYRTTEKKAFFGLLELVFYQNSLKQCFPS 756  
QY 1307 LD-----NPPAAEAASTPAASYVVKVSLIPGGVGAAVVAASNALIPDSGQSESEL 1359  
DB 757 LDTTLFPYKPEPRRAISKPPAGSTKY-----FQTAKARYDFCARD---RSEL 801  
QY 1360 SEVEG 1364  
DB 802 SLKEG 806  
  
RESULT 13  
TRIO\_HUMAN  
ID TRIO\_HUMAN STANDARD; PRT; 3038 AA.  
AC 075962; O13458;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Triple functional domain protein (PDPF interacting protein).  
GN TRIO.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
OX NCBI\_taxid:9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Fibroblast;  
RX Depandt A., Serra-Pages C., Seipel K., O'Brien S., Tang M., Park S.-H.,  
RA Streuli M.;  
RT "The multidomain protein Trio binds the LAR transmembrane tyrosine  
RT phosphatase, contains a protein kinase domain, and has separate rac-  
RT specific and rho-specific guanine nucleotide exchange factor  
RT domains".  
RL Proc. Natl. Acad. Sci. U.S.A. 93:5466-5471(1996).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC Streuli M.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP STRUCTURE BY NMR OF 1227-1407, AND MUTAGENESIS.  
RX MEDLINE-99005194; PubMed-9790533;  
RA Liu X., Wang H., Eberstadt M., Schmuchel A., Olejniczak E.T.,  
RA Meadows R.P., Scherfyanetz J.M., Janovick D.A., Hartlan J.E.,  
RA Harris E.A.S., Staunton D.E., Festik S.W.;  
RT "NMR structure and mutagenesis of the N-terminal Dbl homology domain  
RT of the nucleotide exchange factor Trio".  
RL Cell 95:269-277(1998).  
RN [4]  
RP CHARACTERIZATION.  
RX MEDLINE-99276567; PubMed-10341202;  
RA Seipel K., Medley Q.G., Kedersha N.L., Zhang X.A., O'Brien S.P.,  
RA Serra-Pages C., Hemler M.E., Streuli M.;  
RT "Trio amino-terminal guanine nucleotide exchange factor domain  
RT expression promotes actin cytoskeleton reorganization, cell migration  
RT and anchorage-independent cell growth".  
RL J. Cell Sci. 112:1825-1834(1999).  
RN [5]  
RP FUNCTION: PROMOTES THE EXCHANGE OF GDP BY GTP. TOGETHER WITH  
CC LEUCOCYTE ANTIGEN-RELATED (LAR) PROTEIN, IT COULD PLAY A ROLE IN  
CC COORDINATING CELL-MATRIX AND CYTOSKELETAL REARRANGEMENTS NECESSARY  
CC FOR CELL MIGRATION AND CELL GROWTH.  
CC [5] SUBUNIT: INTERACT TO FORM A COMPLEX WITH LEUCOCYTE ANTIGEN RELATED  
CC PROTEIN.  
CC [5] ALTERNATIVE PRODUCTS: 2 ISOFORMS, 1 (SHOWN HERE) AND 2; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC [5] TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, SKELETAL MUSCLE,  
CC BLADDER, PANCREAS, PLACENTA, LIVER, KIDNEY AND LUNG.  
CC [5] DOMAIN: THE N-TERMINAL DBL/GEF DOMAIN SPECIFICALLY CATALYZES  
CC NUCLEOTIDE EXCHANGE FOR RAC1, LEADING TO THE ACTIVATION OF JUN  
CC KINASE AND THE PRODUCTION OF MEMBRANE RUFILES. THE SECOND DBL/GEF  
CC DOMAIN IS AN EXCHANGE FACTOR FOR RHOA AND INDUCES THE FORMATION OF  
CC STRESS FIBERS.  
CC [5] PTM: SERINE PHOSPHORYLATED.  
CC [5] SIMILARITY: BELONGS TO THE RHO/RAC GEF FAMILY.  
CC [5] SIMILARITY: CONTAINS 2 DBL-HOMOLOGY DOMAINS (DH).  
CC [5] SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.  
CC [5] SIMILARITY: CONTAINS 2 PH DOMAINS.  
CC [5] SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
CC [5] SIMILARITY: CONTAINS 4 SPECTREIN REPEATS.  
CC [5] SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SER/THR  
CC FAMILY OF PROTEIN KINASES.  
CC [5] THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; AF091395; AAC43042.1; -

DR	EMBL: U42390; AAC34245.1; -.		
DR	HSSP: Q63450; 1A06.		
DR	MM: 601893; -.		
DR	InterPro: IPR001251; CRL_TRO.		
DR	InterPro: IPR000719; Euk_pkinase.		
DR	InterPro: IPR003006; Ig_MHC.		
DR	InterPro: IPR003598; Ig_c2.		
DR	InterPro: IPR001849; PH.		
DR	InterPro: IPR000219; RhogEF.		
DR	InterPro: IPR001452; SH3.		
DR	InterPro: IPR002290; Ser_thr_kinase.		
DR	InterPro: IPR002017; Spectrin.		
DR	Pfam: PF00047; Ig; 1.		
DR	Pfam: PF00169; PH; 2.		
DR	Pfam: PF00069; pkinase; 1.		
DR	Pfam: PF00621; RhogEF; 2.		
DR	Pfam: PF00018; SH3; 1.		
DR	Pfam: PF00435; spectrin; 7.		
DR	SMART: SM00408; IgG2; 1.		
DR	SMART: SM00233; PH; 2.		
DR	SMART: SM00325; RhogEF; 2.		
DR	SMART: SM00516; SEC14; 1.		
DR	SMART: SM00326; SH3; 2.		
DR	SMART: SM00150; SPEC; 6.		
DR	SMART: SM00220; S_TKC; 1.		
DR	PROSITE: PS00741; DH_1; FALSE_NEG.		
DR	PROSITE: PS50010; DH_2; 2.		
DR	PROSITE: PS50003; PH_DOMAIN; 2.		
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.		
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.		
DR	PROSITE: PS00108; PROTEIN_KINASE_ST; 1.		
KW	Guanine-nucleotide releasing factor; Phosphorylation; Repeat;		
KW	Transferase; Serine/threonine-protein kinase; ATP-binding;		
KW	Immunoglobulin domain; SH3 domain; Alternative splicing.		
FT	REPEAT	252	359
FT	REPEAT	479	585
FT	REPEAT	819	925
FT	REPEAT	1050	1157
FT	DOMAIN	1233	1408
FT	DOMAIN	1421	1532
FT	DOMAIN	1597	1653
FT	DOMAIN	1910	2086
FT	DOMAIN	2098	2212
FT	DOMAIN	2630	2993
FT	DOMAIN	2737	2972
FT	DOMAIN	656	659
FT	DOMAIN	1786	1791
FT	DOMAIN	2233	2253
FT	DOMAIN	2486	2492
FT	DISULFID	2637	2700
FT	NP_BIND	2743	2751
FT	BINDING	2766	2766
FT	ACT_SITE	2856	2856
FT	VARSPPLIC	2242	2308
FT	VARSPPLIC	2309	3038
FT	MOTAGEN	1240	1240
FT	MOTAGEN	1244	1244
FT	MOTAGEN	1330	1330
FT	MOTAGEN	1367	1367
FT	MOTAGEN	1368	1368
FT	MOTAGEN	1369	1369
FT	MOTAGEN	1371	1371



FT MTAGEN 1372 1372 K->A: LOSS OF NUCLEOTIDE EXCHANGE  
 FT MTAGEN 1375 1375 L->A: 40% DECREASE IN NUCLEOTIDE EXCHANGE  
 FT MTAGEN 1378 1378 ACTIVITY.  
 FT MTAGEN 1379 1379 K->A: NO CHANGE IN NUCLEOTIDE EXCHANGE  
 FT MTAGEN 1379 1379 ACTIVITY.  
 FT MTAGEN 1379 1379 E->A: 30% DECREASE IN NUCLEOTIDE EXCHANGE  
 FT MTAGEN 1379 1379 ACTIVITY.  
 SQ SEQUENCE 3038 AA; 341611 MW; 28620F3B513EB74B CRC64;

Query Match 3.0%; Score 238.5; DB 1; Length 3038;  
 Best Local Similarity 17.6%; Pred. No. 0.0025;  
 Matches 272; Conservative 195; Mismatches 478; Indels 599; Gaps 66;

QY 229 LVSQADSG--LDSGTERPSSISESLMNRNSVLSDPGLDSPOTSVYLARVAQHRR-- 284  
 DB 1189 ISSDSNRSSKSLQDIDIPASIPGSEVKLRDAHLENEKRKSARKKEFIMAEIJOTERKAY 1248  
 QY 285 ---RQSDAALLPLNHQIDOSP---KPLIT--GPEEDYDGFVNNESDITFODLEK 333  
 DB 1249 VRDLRECDMTYLEMNT--SGVEIPIPGIYNKELLIFGNMKEIYE---FHN--NIFLKELEK 1302  
 QY 334 LKSHPA-----YLVFLRYILSQADPGPLLFYLCSEVYQQTNPKDSRSLGKDIW 382  
 DB 1303 YEQLPEDEVGHCFYVWADRFQMYVYCKNKPDSITOLILEHAGSYDEIQOR--HGLANSIS 1360  
 QY 383 NIFLEKNAPL-RVKIPEMLAEIDLRLNNEPRNVLCENQAEVWLEQOENINRYRKRT 441  
 DB 1361 SYLIK---PVOIRTKYQLLKEKEL--LTCCBEGKEITKDGLE--VMLSPKRAKD-----A 1408  
 QY 442 LGISLGVENDLGLDG-----DPL-----REROMAEKQALALDILSKYEED 484  
 DB 1409 MHLMLBGFEDINISQGLILIQESFYQWDPKTLIRKGERHLF--LFEMSLVFSKEVD 1465  
 QY 485 RSAPMDAVNT--FMSNAGITLRESRSSC-----TAE-----KTOSAPDKMWLP 527  
 DB 1466 SSGRSKYLKSKLPTSELGVEHEVGDPCFALWVGRTPTSDNKIVLKASIEKODMIK 1525  
 QY 528 FF-----PKTKOSNSKKEKDALEDKRN----- 553  
 DB 1526 HIREVDEERTIHLKGAKEPHIHTKAPATROKGRDGEDDSOGDDSDPDTIISART 1585  
 QY 554 -----ILRYIGKPKSSSQSISIKGVNRIIQHFNESHQYDVE----- 590  
 DB 1586 SQNTLSDSKLSGGCELIVYIHDFACNSNELTIRGQTVLE-----RPADKDWCLVR 1640  
 QY 591 -----PGTORL-----STGFPEDILLE 607  
 DB 1641 TTDRSPAAGLIVPCGSLCIAHSRSSMEMEGIFNHNKDSLVSNDASPPASVSLQPHMIG 1700  
 QY 608 SDDS-----RSEIRLGRSGSLKGR-----EEMKRSRKANVPRPSDVMD 648  
 DB 1701 AQSPPGKRREONTLRKMLTSPYRLSSGKADGHYKLAHKKRREVRKSDAGSOKDSD 1760  
 QY 649 AAA-----EAARLHQASASSASS----- 666  
 DB 1761 DSAATPODEVEERGRNEGILSSGTLSSSSGMSGCEGEGGADAVLPPLPAIIOHS 1820  
 QY 667 -----LSTRSLENPPPTFP-----KMGKRSIESPULGCTIVYL 701  
 DB 1821 LLOPDSODDKASSRLVAPRPSSETPSAAEIVSAIEELVKSMALEDPRSSILVDOGSSS 1880  
 QY 702 PHLEDDLGOLSDLEPEREVOMOHITGKDVANILTOREIDROVIELFTEASHLRTL 761  
 DB 1881 PSFNPDSNLSSSSPIDEEMERKSSSLK-----RRHVVLDVLETEDDYRDL 1929  
 QY 762 RVLDLIFYQMRKKNEMPREELAR--LFPNLPLEIITHNSWCAMKLRREGPIIDIS 818  
 DB 1930 GYV--VEGYMALMKEDGVDDMKGDKIVFGNIHQIDMHRDF-----LGELEKCLE 1980  
 QY 819 DPHLARDEGARRELDQVAAQFCYSQVALELITKQKREKSFOLDQOEASHQC----- 874

DB 1981 DP-----EKLGSLEFYKH-----ERRLHMYIAYQCNKPKSEHIV 2013  
 QY 875 -----RRLOLRDIYSEMRQRLKRPILLENIITKHGEGSEHEKLCRAR 918  
 DB 2014 SEYIDFEEDLKORLGRHLDLTLLKVPORIMKYOLLDLDPKYSKASDLSELERA- 2072  
 QY 919 DQCEILKFVNEAVKQTEENHRRLEGYQKRLDATALEBASNPAAEFKSLDITRKMHEG 978  
 DB 2073 ---VEVMCIYPRKRCNDMMNNGRLQGF-----GKIYAOG 2103  
 QY 979 PLTWIRISKRTLDLQVLLLEDVLLOROEERLLKCHSK-----TAVGSS--DSKOT 1029  
 DB 2104 K-----LLLOD--FLVLYDODAGILLPRCERBRIFLEQIVIFSEPLDKKK 2147  
 QY 1030 FS-----PVKLKNAV-LIRSVATDKRAFFIITCSLSPQIYELVALTSSDKNT---W 1078  
 DB 2148 FSNPGLFKNSIKVSCLEENENDEKCPAL--TSRTG--DVEVFILHSSSPSVKQTV 2203  
 QY 1079 MELLEAVONATKHPGAAPIPH-----PSPPG---SQ 1108  
 DB 2204 IHEINOLLENOBNFLNLTSPIEYQRNHSGGGGGSGSGSGGGGAGAPSGGSGHSG 2263  
 QY 1109 EPAYOGSTSSRVELNDSEVYHTEKEPKKLPDEGPEBOVODKOLIAQEPVQEBDEELR 1168  
 DB 2264 GPSSCGGAPST-----SRSRPSRIPQ----- 2284  
 QY 1169 TLPAPPSLDGENKGRITRDPVLLALTGPLMGLADALEDEVENLHLLMSLLPGHTV 1228  
 DB 2285 -----PVNHPHVL-----V 2294  
 QY 1229 KTOAGEPEDD-----LTPPSVYVITSHPDGSPGOAPTISDSTRLARPEGSSQPEDG 1283  
 DB 2295 SSAASQAEADKMSGTSTPGPSL-----PPGCAPEAGPSAPSRPPADAGESE 2344  
 QY 1284 VAVSIALHLPPRTSSGVMPSPELDNRPAEASTERPAASYKVR--KVSLLPGG--GVGA 1340  
 DB 2345 REAEPV---PKMK---VLESPP--RKGAAVSSSSDPAKAKARASLTGLTLPGLKPRAGA 2394  
 QY 1341 AKYAG---SNAIPDSGQS---ESELSEVEGAQATGNCFYVSPAGPL-----D 1383  
 DB 2395 ASPLNLSAVPSLGEKPEPPSPSLQ-----GGSFWSSIPASPARPSGFTPPGD 2446  
 QY 1384 SSTEPGTGTPSPSQCHSLPAMPTEPDPYRGVKGCCSSLVRRDV 1427  
 DB 2447 SDSLQROTTPR-----HAAPGKRTD-----RMSCTSSASEQSV 2478

RESULT 14  
 ECT2\_MOUSE  
 ID ECT2\_MOUSE STANDARD; PRT; 738 AA.  
 AC 007139;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 01-MAR-2002 (Rel. 41, Last annotation update)  
 DE ECT2 protein (Epithelial cell transforming sequence 2 oncogene).  
 GN ECT2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=93218723; Pubmed=8464478;  
 RA Miki T., Smith C.L., Long J.E., Eva A., Fleming T.P.;  
 RT "Oncogene ect2 is related to regulators of small GTP-binding  
 RT proteins";  
 RL Nature 362:462-465(1993).  
 RN [2]  
 RN ERRATUM.  
 RA Miki T., Smith C.L., Long J.E., Eva A., Fleming T.P.;  
 RL Nature 364:737-737(1993)  
 CC -!- FUNCTION: BINDS HIGHLY SPECIFICALLY TO RHQ, RHOC AND RAC  
 CC PROTEINS, BUT DOES NOT APPEAR TO CATALYZE GUANINE NUCLEOTIDE

CC EXCHANGE. TRUNCATION OF THE N-TERMINUS PUTS THE PROTEIN IN AN  
 CC ACTIVATED STATE.  
 CC -I- TISSUE SPECIFICITY: HIGHEST EXPRESSION IN TESTIS. ALSO DETECTABLE  
 CC IN KIDNEY, LIVER AND SPLEEN.  
 CC -I- SIMILARITY: CONTAINS 2 BRCT DOMAINS.  
 CC -I- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
 CC -I- SIMILARITY: CONTAINS 1 PH DOMAIN.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC -----  
 CC EMBL: L11316; AAA37536.1; -  
 CC MGD: MGI:95281; Ect2.  
 CC DR InterPro: IPR001357; BRCT.  
 CC DR InterPro: IPR001331; GDS\_CDC24.  
 CC DR InterPro: IPR001849; PH.  
 CC DR InterPro: IPR000219; RhogEF.  
 CC DR Pfam: PF00533; BRCT; 2.  
 CC DR Pfam: PF00621; RhogEF; 1.  
 CC DR SMART: SM00292; BRCT; 2.  
 CC DR SMART: SM00233; PH; 1.  
 CC DR SMART: SM00325; RhogEF; 1.  
 CC DR PROSITE: PS50172; BRCT; 2.  
 CC DR PROSITE: PS50010; DH; 2; 1.  
 CC DR PROSITE: PS00741; DH; 1; 1.  
 CC DR PROSITE: PS50003; PH DOMAIN; 1.  
 CC KW Guanine-nucleotide releasing factor; Repeat; Proto-oncogene.  
 CC FT DOMAIN 1 76 BRCT 1.  
 CC FT DOMAIN 91 179 BRCT 2.  
 CC FT DOMAIN 277 466 DH.  
 CC FT DOMAIN 500 619 PH.  
 CC SEQUENCE 738 AA; 83685 MW; 3D270CBF1570C681 CRC64;  
 Query Match 3.0%; Score 237; DB 1; Length 738;  
 Best Local Similarity 19.4%; Pred. No. 0.00046;  
 Matches 133; Conservative 118; Mismatches 259; Indels 174; Gaps 25;

DB 546 TRPPASLKIHIMPLSIOIKKVLDIRTEDCHNAFALLVRPPEQANVLLSPKOTSEELPK 605  
 QY 1076 NIMWELLEAVONATKHPGAPPIHPSPGSGEPAYOGSTSSRVEINDSEVYHTEKEPK 1135  
 DB 606 ESKLMLKCHRVANTICKAENL-MYVADPE-----FEVNTKDM----- 644  
 QY 1136 KLEPGGPEQRVODKOLIOGEPVOEDEBEL--TLPPAPSLDGENGIRTRDPVLLA 1193  
 DB 645 -----DSTLSRASRAIKTKSKVYRAFESKTP-----KRALR-----MA 679  
 QY 1194 LFGPLMEG-----LADAALEDVENLRHLIWSLPGHVTQTOAGEPEDDLTPPS 1245  
 DB 680 LSSSHSEGRSPSSSKLAVSLRSTSSL-----AQPSPSLVSLPS 721  
 QY 1246 VVSTSHPMDPGSPGQAPTIIDST 1269  
 DB 722 FFERASH-----TLRST 734  
 RESULT 15  
 VAV\_HUMAN STANDARD; PRT; 845 AA.  
 ID VAV\_HUMAN  
 AC P15498; Q15860;  
 DT 01-APR-1990 (Rel. 14; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 01-MAR-2002 (Rel. 41; Last annotation update)  
 DE VAV proto-oncogene.  
 GN VAV1 OR VAV.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Denkiner D.J., Borges C.R., Shaw C.L., Cushman A.M., Kawahara R.S.;  
 RT "Transcriptional regulation of the vav proto-oncogene."  
 RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE OF 62-845 FROM N.A.  
 RX MEDLINE=90005432; PubMed=2477241;  
 RA Katay S., Martin-Zanca D., Barbacid M.;  
 RT "vav, a novel human oncogene derived from a locus ubiquitously  
 RT expressed in hematopoietic cells."  
 RL EMBO J. 8:2283-2290(1989).  
 RN [3]  
 RP SEQUENCE OF 1-61 FROM N.A.  
 RX MEDLINE=91172176; PubMed=2005887;  
 RA Katay S., Cleveland J.L., Heslop H.E., Pulido D.;  
 RT "Loss of the amino-terminal helix-loop-helix domain of the vav proto-  
 RT oncogene activates its transforming potential."  
 RL Mol. Cell. Biol. 11:1912-1920(1991).  
 RN [4]  
 RP SEQUENCE OF 299-837 FROM N.A.  
 RA Romero F.;  
 RL Submitted (JAN-1995) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RP SEQUENCE OF 299-334 FROM N.A.  
 RX MEDLINE=96038895; PubMed=7478592;  
 RA Ramos-Morales F., Romero F., Schweighoffer F., Bismuth G., Camonis J.,  
 RA Tortolero M., Fischer S.;  
 RT "The proline-rich region of Vav binds to Grb2 and Grb3-3."  
 RL Oncogene 11:1665-1669(1995).  
 RN [6]  
 RP SIMILARITY TO CDC24 FAMILY.  
 RX MEDLINE=92228488; PubMed=1565462;  
 RA Adams J.M., Houston H., Allen J., Lints T., Harvey R.;  
 RT "The hematopoietically expressed vav proto-oncogene shares homology  
 RT with the dbl GTP-GDP exchange factor, the bcr gene and a yeast gene  
 RT (CDC24) involved in cytoskeletal organization."  
 RL Oncogene 7:611-618(1992).  
 CC -I- FUNCTION: PROBABLE EXCHANGE FACTOR FOR A SMALL RAS-LIKE GTP-  
 CC BINDING PROTEIN. CAN BE ACTIVATED BY TRUNCATION OF THE N-TERMINUS.  
 CC -I- TISSUE SPECIFICITY: WIDELY EXPRESSED IN HEMATOPOIETIC CELLS BUT

CC NOT IN OTHER CELL TYPES.  
CC -1 SIMILARITY: CONTAINS 1 CALPONIN-HOMOLOGY (CH) DOMAIN.  
CC -1 SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).  
CC -1 SIMILARITY: CONTAINS 1 PH DOMAIN.  
CC -1 SIMILARITY: CONTAINS 1 ZINC-DEPENDENT PHORBOL-ESTER AND DAG BINDING DOMAIN.  
CC -1 SIMILARITY: CONTAINS 1 SH2 DOMAIN.  
CC -1 SIMILARITY: CONTAINS 2 SH3 DOMAINS.  
CC -1 CAUTION: REF 2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 322 TO 355 DUE TO A FRAMESHIFT.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
DR EMBL; AF030227; AAC25011.1; -  
DR EMBL; AF030201; AAC25011.1; JOINED.  
DR EMBL; AF030202; AAC25011.1; JOINED.  
DR EMBL; AF030203; AAC25011.1; JOINED.  
DR EMBL; AF030204; AAC25011.1; JOINED.  
DR EMBL; AF030205; AAC25011.1; JOINED.  
DR EMBL; AF030206; AAC25011.1; JOINED.  
DR EMBL; AF030207; AAC25011.1; JOINED.  
DR EMBL; AF030208; AAC25011.1; JOINED.  
DR EMBL; AF030209; AAC25011.1; JOINED.  
DR EMBL; AF030210; AAC25011.1; JOINED.  
DR EMBL; AF030211; AAC25011.1; JOINED.  
DR EMBL; AF030212; AAC25011.1; JOINED.  
DR EMBL; AF030213; AAC25011.1; JOINED.  
DR EMBL; AF030214; AAC25011.1; JOINED.  
DR EMBL; AF030215; AAC25011.1; JOINED.  
DR EMBL; AF030216; AAC25011.1; JOINED.  
DR EMBL; AF030217; AAC25011.1; JOINED.  
DR EMBL; AF030218; AAC25011.1; JOINED.  
DR EMBL; AF030219; AAC25011.1; JOINED.  
DR EMBL; AF030220; AAC25011.1; JOINED.  
DR EMBL; AF030221; AAC25011.1; JOINED.  
DR EMBL; AF030222; AAC25011.1; JOINED.  
DR EMBL; AF030223; AAC25011.1; JOINED.  
DR EMBL; AF030224; AAC25011.1; JOINED.  
DR EMBL; AF030225; AAC25011.1; JOINED.  
DR EMBL; AF030226; AAC25011.1; JOINED.  
DR EMBL; X16316; CAA34383.1; ALT\_FRAME.  
DR EMBL; M59834; AAG3267.1; -  
DR EMBL; X83931; CAA58783.1; -  
DR PIR; S05382; TVHUV.  
DR HSSP; P29354; IGR1.  
DR TRANSFAC; T00880; -  
DR MIM; 164875; -  
DR InterPro; IPR003247; CH\_type.  
DR InterPro; IPR001715; Calponin\_hom.  
DR InterPro; IPR002219; DAG\_PE-bind.  
DR InterPro; IPR001331; GDS\_CDC24.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR000219; RhogEF.  
DR InterPro; IPR000960; SH2.  
DR InterPro; IPR001452; SH3.  
DR InterPro; IPR003096; SM22\_calponin.  
DR Pfam; PF00307; CH; 1.  
DR Pfam; PF00130; DAG\_PE-bind; 1.  
DR Pfam; PF00169; PH; 1.  
DR Pfam; PF00621; RhogEF; 1.  
DR Pfam; PF00017; SH2; 1.  
DR Pfam; PF00018; SH3; 2.  
DR PRINTS; PR00401; SH2DOMAIN.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR PRINTS; PR00888; SM22CALPONIN.  
DR ProDom; PD001527; CH\_type; 1.

DR SMART; SM00109; C1; 1.  
DR SMART; SM00033; CH; 1.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00325; RhogEF; 1.  
DR SMART; SM00252; SH2; 1.  
DR SMART; SM00326; SH3; 2.  
DR PROSITE; PS50021; CH; 1.  
DR PROSITE; PS50047; DAG\_PE\_BIND\_DOM\_1; 1.  
DR PROSITE; PS50081; DAG\_PE\_BIND\_DOM\_2; 1.  
DR PROSITE; PS50010; DH\_2; 1.  
DR PROSITE; PS00741; DH\_1; 1.  
DR PROSITE; PS50003; PH\_DOMAIN; 1.  
DR PROSITE; PS50001; SH2; 1.  
DR PROSITE; PS50002; SH3; 2.  
KW Proto-oncogene; Phorbol-ester binding; Zinc; SH2 domain; SH3 domain;  
KW Guanine-nucleotide releasing factor; Repeat.  
FT DOMAIN 1 119 CH.  
FT DOMAIN 194 373 DH.  
FT DOMAIN 402 504 PH.  
FT DOMAIN 516 564 PHORBOL-ESTER AND DAG BINDING.  
FT DOMAIN 617 660 SH3 1.  
FT DOMAIN 671 765 SH2.  
FT DOMAIN 782 842 SH3 2.  
FT CONFLICT 264 264 SH3 2.  
FT CONFLICT 718 718 A -> P (IN REF. 2).  
FT CONFLICT 718 718 I -> TV (IN REF. 2).  
SQ SEQUENCE 845 AA; 98313 MW; AC3BC9736FD2F138 CRC64;

Query Match 3.0%; Score 233.5; DB 1; Length 845;  
Best Local Similarity 21.7%; Pred. No. 0.00079;

Matches 151; Conservative 93; Mismatches 267; Indels 185; Gaps 30;

QY 711 QLSDEPEPE-----YOMOHVGVKDVANLTQRE-----IDREVINEFVTE 754  
DB 148 QIDDIYEDDEDLYDCVBN-EEAGDEIYEDIMRSEVSNPPKTEYDKRCCCLREIQOTE 206  
QY 755 ASHLRTLVLDLIFYOMRKRENIMPREELARLPNDPELEIHSWCEAMKRLREGPTI 814  
DB 207 EKYTDIGSIQGHFLKPLQR--FLKPDQIEIIFINIEDLLRVTHFLKEKE----- 256  
QY 815 RQISDMALARFDPAREELQO-----VAAGCS-QVSVALILIRKOKRESFOL 863  
DB 257 -----ALGTPGANLYOVPIKYKEREFLVYGRCSOVESASKHLDVYAAARD-YOM 306  
QY 864 FMOAESHPQCRRLQRLDLIVSEMRUTKYPLLENIKHEGTSSEHEKLCRAPQCE 923  
DB 307 KLECCSGORANNGRFTLRDLIMVPMQVLYKHLLOELVHTQ-EEAKKEKRLALDAMD 365  
QY 924 ILKFNVAVKQTEENRRHLEGYQ--KRLDATALERASNPPLAEFKSLDITRRMIEGPI 980  
DB 366 IAGCVNEVKRDNETLRQITNFQSLIENLQSLAHYGRPKIDGLK--ITS----- 413  
QY 981 TWRISKDTLDVLLLEDLVYLLQROEERLLK-----CHS-----KTAVGSSSKOTFSP 1032  
DB 414 ---VERRSKMDRYAFLDLAALLCKRRGDSYDLKDFVNLSHFQVRDSSGDRNK----- 465  
QY 1033 VLKLNVLILRSVATDRAEPIICTSELGPOIYELVALTSSDKNIMMLEEPAVNATKH 1092  
DB 466 -----KSHHFLIIEDOG-AQGTLEFKRELKMKMEQFEALISNI--Y 507  
QY 1093 PGAAPPIPIHSPSPSQEPA-----YQGSTSRVEINDESEYHTEKEPKKLP 1138  
DB 508 PENATANGHDQFMFSFEETTSCKACQMLRGTFYQGYRCHRCAS----AHKRCGLGRVP 563  
QY 1139 EG-----EGP-----EQRVQDKQLAAGEPVQOEDEDEELKRLPRAPSLGGENGIRT 1186  
DB 564 CGRHGODFPGTMKDKLHRAADQKKRNEGLPRMEVFQOEYGLRP--PGAIG----- 614  
QY 1187 RQVVLALGPIIMBGLAQA--LEDVENLRLLILMSLIPGTVYQQAAGEPEDDLPFT 1243  
DB 615 --PELRNPGDIVELTAKAEDEQNMWEGKNTSTNEIGW--PPCARVAPYVHGPPQD----- 665  
QY 1244 PSVVSITSHPWPGSPGOAFTISDSTRLARPEGSGEGEDVAVSSIAHLPPRTRSSGVD 1303

Db 666 -----LSVHLMWYAG-----PMERAGAESILAN-----RSDGTF- 693  
Oy 1304 SPELDRNPAAEASTEPAAASYKV-VRKVSLPBGGV 1338  
Db 694 ---LVROKVKDPAEFAISIKYNEVEVKHKIKIMTAEG 726

Search completed: September 18, 2002, 10:43:27  
Job time: 328 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 10:35:58 ; Search time 31.34 Seconds  
(without alignments)  
4681.826 Million cell updates/sec

Title: US-09-695-795-4

Perfect score: 7883

Sequence: 1 MSIRLPHSIDRSASRKQSHL.....EGSDIALEDSATDTAVSPGP 1527

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: PIR\_71:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	922.5	11.7	2559	2 T09144	probable guanine n
2	672	8.5	979	2 A89633	protein F136.6 [1
3	521	6.6	1693	2 T30867	Rho-guanine nucleo
4	462	5.9	573	2 T49342	hypothetical oncog
5	428	5.4	424	2 T38434	P47 LBC oncogene -
6	318	4.0	1805	2 A34736	nestin - rat
7	297.5	3.8	1737	2 A59235	unconventional myo
8	297	3.8	1804	2 T34518	nestin - golden ha
9	262.5	3.3	1356	2 S51389	ROM2 protein - yea
10	261	3.3	1158	2 T50454	probable rho1 GDP-
11	260.5	3.3	1334	2 T41524	rho1 gdp-gtp excha
12	256.5	3.3	1046	2 T34566	hypothetical prote
13	255.5	3.2	5327	2 T13564	microtubule-associ
14	247.5	3.1	1618	2 S21424	nestin - human
15	239.5	3.0	1045	2 JC5795	CDEP protein - hum
16	239	3.0	2101	2 A42184	nuclear mitotic ap
17	237	3.0	738	2 S32372	transforming prote
18	231.5	2.9	460	2 G01210	guanine nucleotide
19	229	2.9	2845	2 T49505	adenomatous polypo
20	228.5	2.9	1829	2 T26135	hypothetical prote
21	227	2.9	878	2 T51940	gene VAN2 protein
22	227	2.9	2774	2 A43359	microtubule-associ
23	226	2.9	1591	2 A54146	invasion-inducing
24	224.5	2.8	648	2 T33339	hypothetical prote
25	223.5	2.8	1155	2 S64365	GDP/GTP exchange p
26	222	2.8	2954	2 T14156	kinesin-related pr
27	221.5	2.8	519	2 T38402	guanine nucleotide
28	221	2.8	1252	2 T14272	coractin-binding
29	220.5	2.8	1621	2 A82255	hypothetical prote

30	220.5	2.8	1684	2 JW0057	gravin - human
31	219	2.8	2044	2 T13704	still life protein
32	219	2.8	2064	2 T13707	still life protein
33	215	2.7	875	2 T19678	hypothetical prote
34	213.5	2.7	961	2 A55380	faciogenital dyspl
35	213.5	2.7	1547	2 T28657	blackjack protein,
36	212	2.7	3938	2 T42761	Bassoon protein -
37	211.5	2.7	753	2 S62411	probable guanine n
38	211.5	2.7	1427	2 S22695	restin - human
39	211.5	2.7	1791	2 T02345	hypothetical prote
40	209	2.7	1131	2 T15617	hypothetical prote
41	209	2.7	1275	2 T41523	hypothetical rho1
42	207.5	2.6	3942	2 T42730	Bassoon protein -
43	206.5	2.6	2442	2 T08621	centrosome associa
44	206.5	2.6	2715	2 T13049	eyelid - fruit fly
45	204.5	2.6	862	2 T34342	hypothetical prote

## ALIGNMENTS

## RESULT 1

T09144

probable guanine nucleotide exchange factor RHOGEF2 - fruit fly (Drosophila melanogas

N:Alternate names: Shar pei/DRHOGEF2

C:Species: Drosophila melanogaster

C>Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 17-Nov-2000

C/Accession: T09144; T09223

R:Hecker, U.; Perrimon, N.

submitted to the EMBL Data Library, October 1997

A:Reference number: Z16586

A:Accession: T09144

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 12559 <HAES>

A:Cross-references: EMBL:AF031930; NID:g2687355; PID:g2687356

R:Barrett, K.; Leptin, M.; Settlement, J.

Cell 91, 905-915, 1997

A>Title: The Rho GTPase and a putative RHOGEF mediate a signaling pathway for the cel

A:Reference number: Z16618; MID:98088790

A:Accession: T09223

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-676, 'L', 678-837, 'L', 839-889, 'S', 891-1243, 'D', 1245-1358, 'E', 1360-1368, 'R

A:Cross-references: EMBL:AF032870; NID:g2760367; PIDN:ACC38820.1; PID:g2760368

C:Genetics:

A:Gene: rhogEF2

A:Cross-references: FlyBase:FBgn0023172

A:Map position: 2; 53Fl-2

A>Note: orchestrates cell shape changes during gastrulation

C:Function:

A:Superfamily: mediates actin rearrangements required for cell shape changes during g

C:Keywords: signal transduction; embryo; GTP exchange

F:1151-1200/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 11.7%; Score 922.5; DB 2; Length 2559;

Best Local Similarity 20.8%; Pred. No. 2.8e-37;

Matches 479; Conservative 279; Mismatches 654; Indels 895; Gaps 82;

QY	17	OSHLSPPIASWLSLSLSLGDSTPRTSPSHRQPSPTSEFTAGLVQRC-----VII 67
DB	204	QSNNSNPVLOAPERSLINTPLSLSGHDESTTPATTPSTPSLALPKNFQYITLTV 263
QY	68	QKDQHFEGFTVSGDRIVLVQSVAPGAAKAGVEGRDRIIKVNGTWVFNSSHLEVVKIK 127
DB	264	RKDSNGYGMKVSQDNPNFVSEVYPGGAELIAGVADMLIRVNGHEVRLEKHPVYGLIK 323
QY	128	SGAVYALTLTIGSS---PSPGV-----SGLOQNPVAVGLVRY----- 161
DB	324	ASTTVELAVKRSSOKLTPRSSVSVPSTPILSGRQRTATISGPQPVDSIKRREMETTKIQ 383

```

QY 162 -----NP----- 163
Db 384 TLQKMLEKLNLERLKSDONNPSYKLSBANIFKLEOLHOGVADAPTUKLOAAAGNKN 443
QY 164 ---LTP-----PPP----- 170
Db 444 TALLPNOIOHLSASATHSNQOFLHHHHNNNNYPPQOQAPASTPAFLSLPRSLSS 503
QY 171 ----- 170
Db 504 LSLGTRKNTKENDLTSSPFGLTDTDLQOORMSHOAESQSMHONTSTPTSOQFFHPHQ 563
QY 171 -----PPPLP---PPQ-----HTTG 182
Db 564 OOHREKETGPTSGKNNKFLISRLIEDYPPPLPQRNPPROLNLDKNGNASPGSHLYA 623
QY 183 P-----KPIODPEYOKHATOILMNM 202
Db 624 PVSDDLDRATSPOLNRSQOOLPRSTDNPSNMAKSKRSKITKALSDP---KMSOTKFLQW 680
QY 203 ----- 202
Db 681 ESASAGAAGSIEVDGPPPLPPLPGMWTEDMSRSGCONLAQPNVGTAFNYPLVSTT 740
QY 203 -----LROBEELDIL-----PCGERSQR----- 223
Db 741 TAVQNDNLNIAEPPLSQRPNIYOOLQOQOQOHOMSGGATGALGOTPILGKNNKRRVGS 800
QY 224 -----TCERLSVDSQEOADSGLDSTGERPSESILM-----NRNS 259
Db 801 SPDNMHRPHRDRITKTTSQSWELVEKDGESS--PRGTPPPYLSSIMTYLEDPENNRGA 859
QY 260 VUSDGP--LDSPQ-----TSPVILARYAQH--HRRQSDALLPLNHOGLDOSPKPLII 309
Db 860 AAGPQGVFTESHQFTPMAGASSPPISLSHNMHMAOSND-----TOKEIIS 906
QY 310 GREEDD-----PGYNNESDILFODEKLKSHPAVLVYRLYILSOADPGPLFLY 361
Db 907 MEDENSDLDEPFLDENGPPNNLRLLEAE-----NVTFLATPLNTYISISDAPLFLY 960
QY 362 CSEVYOOTNPKDSRSLGKDIWNIFLEKNAPLR--VKIPEMLQAEIDLRLNNEP---RN 416
Db 961 ITELYEGTSKDMRKMAVEIHSFVLVAPRLSWYRODESAREVDNVULOEDYKVELRT 1020
QY 417 VLCEAOAVMLETOEOINDYRSKRTIGLSLYGEND--LLGLDGPDLREKMAEKOL--- 471
Db 1021 VELRSKRAKADLISEQRLREFQOKRTAGLTGYGPTDKLAEAKTDLRE--QIIDKYLMPN 1079
QY 472 -----AALGDIKSYEEDRSAPMDEFA--VNTFMSH----- 499
Db 1080 LHALIIDEKSPPEEDYRKVALCSALSTVYIRIENTPRPPSSIVERYHNHVSRKSKSR 1139
QY 500 -----AGIR---LRE-----SRSSCTAETKOS 518
Db 1140 MGNRKRMNVRGHPVLVROYEVTCHNCQOTIIMGVSPQGHCTDKLNIHQOSKVVDES 1199
QY 519 APDXOKWLPFPPTKROSSKKEK-----DALDKKR----- 551
Db 1200 CQEP-----LPOAKRLAHNDKISKFMGKIRPRTSDVIGEKRSRODEELNELLPTDRQO 1253
QY 552 -----NPILYIKPKSSOSIKPGVNRNIIIOHFENSHQYDVEPG----- 592
Db 1254 ASIVRPSDRRPRANISISNGTSCNTSGL---NTTDLOSSHGSCADSTIPGGGAGC 1310
QY 593 TORLST-----GSFPE---DLLESDDR-----SEIRLGRS 620
Db 1311 NMDLSTSVASTPTSTGSVAAGLSAPAEALNALDVTYDKEARREYRSHQPHKKSAPVSNMS 1370
QY 621 GSLKGEKEKRSKRAEN-----VRPRSVDVMDAAAEARLHQSASSASSLS----- 668
Db 1371 ESYKERLSMKNRNRNRSRKTSDPSLSSRP--NDEQLDLGLSNATYVGSNSLSAGGTESP 1429
QY 669 TSLERPTPP-----FTPKMGRRSIESPNL-----GFCOTVILPHILEDLQGLS-- 713

```

```

Db 1430 STSMHEFAAPGAAGVOYPPMGLNONOHPHLLIOOHNOQYCCOODSFQAGLAGAASSAAS 1489
QY 714 -----DLEPEYQO---NMOHTGKDVANULTOREIDREVINELFVY 753
Db 1490 NSSFMNAGHPRLVYARKTLESEDEDDVENDADSSMVAABVSAALTDAEKKROEIIINEIYOT 1549
QY 754 EASHLRTLVLDLIFYOORKNENLMPREELARLF--PNIPELIEIHSNWCAMKKLR--EEG 811
Db 1550 ERNHVATTKLDRFLFLPYLESGLISODHLLFPFALLSTREIGHAFOSIKORRIEHN 1609
QY 812 PIIRIDSPMLARFPGPARAEELQOYAAOCFSQVALLEIRKOKRESFOLPMQAESEH 871
Db 1610 HAVNTITGDLAMFQGSQSVLCEFPACQAOQALALELKKRKNDEMLQRLKKSSEH 1669
QY 872 POCSRLOLRDLIVSEMQRLTKYPLLENIITKHT-----EGSTSEHKLKRAPDQCEIILKF 927
Db 1670 KACRRELKDLPLTVLORTKTYPLFENIYKTVRLRPENTTEAEALQRAVESKRILVE 1729
QY 928 VNEAVKQTENRHRLSEGYOKRLDATALERASNPLAEFKSLDITTKMHEGLTRISKD 987
Db 1730 VNQAVRTADAHKLOINQKRLDRSSYDK-----EEFKLDLTQHHLLJHDGMLT--IKKN 1781
QY 988 KTLDLQVLLLEDLVYLLQOREERLLK--CHSKTAVSSDSKOTFSPLYKLNA--VLIRSYA 1045
Db 1782 PSVOLHGLLFEMNIYLLTQDDKYIKNLHTPLST---TNKPRPIMSIDADTLIRQEA 1837
QY 1046 TKRAFFIITSELGPQIYELVALTSSDKNIMELLEBAVONATKH----- 1092
Db 1838 ADKNSFFLI---KMTSOMLELRAFPSSSECKTWFKFHSVDVAAROSKNRSKNASNHDTSI 1894
QY 1093 ---PGAPRIPIHSPGSOE-----PAYGSTSS 1118
Db 1895 SDPALAIP--HSNTKESTLESTDTVOPLAATLTTTPRLAPMLPIATYTPRABATNNSNVS 1953
QY 1119 -----RVEINDSEVYHNEKPKKLEPGRCPEORVODKOLI--AOGEPYO-----EED 1163
Db 1954 SILTVGLRNRPQDARATASEDADYVNTPKPRSSONENVRNTRMSRSGEPLOKXSANGTEAN 2013
QY 1164 EEBELR---TLPRAPPSLDENR-----GIRTRDYLAL-----TGPLMBG--LA 1204
Db 2014 DVTLRHSQSTRSESVRGSTGEERNSTYGVNGNSKRDASIVCSNNSNTRTLMOQSPLY 2073
QY 1205 D-ALEDEVNHLHLLMSLPGHTVKTQAAGERD-----DLTPP-----SVYSI 1249
Db 2074 DPTAIOVSTSPAHNTAEVLTLPGEKLRLDASTIRNDLEKOKITICDIFRLPVHTQOYDI 2133
QY 1250 TSHPMDPGSPGAPRTISDSTRLARPEGSQ-----PEGEDVAVSL---AH- 1291
Db 2134 AMMP-----EAPKDSADIALAAVDOIQTLTKMLNEYHNVPEQGVSAVSTAVCGHCH 2186
QY 1292 -----LPPRTSSGVWDSPELDNRNPALEAASPERAASYVVKRVSLPRGGVGAAY 1343
Db 2187 KEKRLKKVAPSSSFSSPPRLPPRNQOHAQAQIPPS---PLMPRLQTLDDEN----- 2238
QY 1344 AGSNAIIPDSGOSELSSEVEGAQAATGNCFYVMPAPPLDSSTTEPTGPSPSQCHSLPA 1403
Db 2239 ---AIHEDDDGCEIDELAR-----LPAIRPKPHRRT--PLAP----- 2272
QY 1404 WPTEPQYRGVGGCCSLVRBDVYIHTI-----EQLTIKLHRL- 1444
Db 2273 FNTERTSQSV---IDASKRQSTDAVPEGLLEOEPLLEDKTEIYGEDNEVXTSPDKLS 2328
QY 1445 ---KDMELAHMELKSLGSESSGGTTPVGSFHTEAARTDVSLSPPAKALASQONOQ 1500
Db 2329 ESCNEERQCEVEDITKEVADPTTSKNEAASVDELPSQSRKEIKTEAMNSKVAADKEDNE 2388
QY 1501 EOGSCPEEGSDIALEDSATDTAVSPGP 1527
Db 2389 E---TIEEGV--ASTYDSSYTQTSPTESP 2411

```

RESULT 2

A89633  
protein F13B6.6 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Nov-2001  
C:Accession: A89633  
R:anonymus, The C. elegans Sequencing Consortium.  
S:Science 282, 2012-2018, 1998  
A:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; MUID:99069613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/projects/C\_eleg  
A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: A89633  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-979 <STO>  
A:Cross-references: GB:chr\_X; PIDN:CAA92119.1; PID:g3875839; GSPDB:GN00028; CESP:F13B6.6  
A:Note: similar to Tiam-1 like protein  
C:Genetics:  
A:Gene: F13B6.6  
A:Map position: X

Query Match 8.5%; Score 672; DB 2; Length 979;  
Best Local Similarity 27.7%; Pred. No. 1.7e-25;  
Matches 223; Conservative 122; Mismatches 270; Indels 190; Gaps 27;  
Oy 415 RVVLEAGDAVWLLEIOEINDYRSKRTGLGSLYGENDLLGLDGDPLERQMAEKQLAAL 474  
Db 15 KRIFPGQRAVTIDINMILNDFRQKKQLG-----GOVGEQ----- 49  
Oy 475 GDILSKYEEDRSAPMDFAVNTFMHAGIRLRSSS-CTAEKTOSAPDKDWLPFPPT- 532  
Db 50 --MFRCEQCSNSTDF-----DSCETRTQSIISLATVTKISAGA-----LQIHPTAS 95  
Oy 533 -----KKOSSNKKKEDALEDKRNPILRKIGKPKSSOSIKGNVNTIQQ----- 578  
Db 96 LKIVLGTSSSSNEKMLD-----KFPQFARADPPKSKRL-PANKIKRIYKVGHSFS 148  
Oy 579 -----HFENSHOYDVEPGTQRTSTGSPFEDLSSRSSE 614  
Db 149 INSVTVHYCYOCRAINGMQPMYIFCSCNDYKVNPHOTSALTACYF---VTQSKQSK 205  
Oy 615 IRL-----GRGSLKGREMKRSRAE-NVPRPSDV-----DMDAA-----EAA 654  
Db 206 SRLSLGIRSDANDEDEGSNAHBEQPNVKSTSDSGIGEHNMGRGLVSRSHSMRYAVA 265  
Oy 655 RLHQSASSASLSRSLNPPPP-----FTRPK-----MGRRSIES 690  
Db 266 TLPOSLSDEKVISPSKRDRSVTPSMORSTGYDLTPADEIDESDIGRKSXYLERRSLES 325  
Oy 691 PN-----LGFCTDVLILPHLEDDL-----GOLSDLEPEPEYQNM 724  
Db 326 SSRMAIDQSYSAASCSHGGSVAMDDICVRRTTTFQFSRPMGD-SDEIETEARPL 384  
Oy 725 QHTYQKVAVANLTQREIDROEVINELFVTEASHLFTLRVLDLIFYQRMKKNLMPREELA 784  
Db 385 EQLGMDVIRHLKPE-----KKOLFETERTHVANKILTVHYEKKPIVTSKIYTEELAN 438  
Oy 785 RLPLPLPELIEHNSWCCEAMK-----KLRECGPIIRIDISDPMLARFGCPAREELQ 835  
Db 439 LFLANLEELMLHKSMSAMDAEVEKMSAPPRVNGIYGIVLMEIMFGAEENLMR 498  
Oy 836 VAAQFCYSVALELIIRT--KQKRESRFOLFMQEAESHPOCRRLQRLDLIYSEMQRLTKY 893  
Db 499 VTATFCQHQHQALEFLRTRCKREKDDAFVRFLAEESNPVCRTQOLKMIIVEMQRLVKY 558  
Oy 894 PLLENIKHTREGGSEHEKLCRAPDOCRELTKFVNEAVKOTENRHRLEGYQKLDATAL 953  
Db 559 PLLLETKAKYTEPSEENCLRTVASAKRTLSAVNTAKRAENMRRLIEELQKRTDSPF 618  
Oy 954 ERASNPPLAAEFKSLDTTRKMIHGGPLTWRTSKDCTLQVLLLEDLVVLQROEE--RL 1011  
Db 619 DKRP--VGHDTYNTMLTKFRLLVHDGPILCRPNRNGKMIELHAVLLENNMLVLTFTKNSDGNKL 676

Oy 1012 LKCHSKTAVGSSDSKOT-FSPVLKLNALVIRSVATDRAFFICTSELPQIYELVAL 1070  
Db 677 VLK-----ALEPSKETRMSPIILPLAPLKEKANDKRAFFLIPNSQYG-AQIYELVAG 728  
Oy 1071 TSSDKNIMELLEBAVONATKHPGA 1095  
Db 729 TATERKTWFKLMGDQIMSEKKNQAA 753  
RESULT 3  
Rho-guanine nucleotide exchange factor - mouse  
T30867  
C:Species: Mus musculus (house mouse)  
C:Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 02-Sep-2000  
C:Accession: T30867  
R:Gebbluk, M.F.B.G.; Poland, M.; Kranenburg, O.; VanHorck, F.P.G.; Moolenaar, W.H.  
submitted to the EMBL Data Library, October 1996  
A:Reference number: 220911  
A:Accession: T30867  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: mRNA  
A:Residues: 1-1693 <GEB>  
A:Cross-references: EMBL:U73199; NID:g1657834; PID:g1657835; PIDN:AAB18197.1  
A:Experimental source: brain  
C:Superfamily: protein kinase C zinc-binding repeat homology  
F:652-698/Domain: protein kinase C zinc-binding repeat homology <KZM>

Query Match 6.6%; Score 521; DB 2; Length 1693;  
Best Local Similarity 20.2%; Pred. No. 1.1e-17;  
Matches 290; Conservative 218; Mismatches 496; Indels 430; Gaps 55;  
Oy 205 QEEELDDILPPCGEPTQRTCEGRLSVDSQEAQSDGDSGTERPFSISELSMNRKSVSDP 264  
Db 283 EQEAKTEKATWPSG-----AAETEVEVNLNLSG--RSPSEEE--DAKSISQ- 326  
Oy 265 GLDSPQTSPIVLARVAQHRRQSGDAALLPLNHQIGDS--PKPLIIGPEEDY----DP 317  
Db 327 -VDPP-----SEH---EDQDLALDRSPFGKLKSKHVPASLAAQGLSDVLANGDE 372  
Oy 318 GYFN-----NESDIIIFDLEKLKSHPAVLVFLRYLTSQADPGPLLYLCEVYQOTN 370  
Db 373 VYANCMVLDQVGDIDIVINIEGLSTH-----TS 401  
Oy 371 PKDSRLSKDQWNTFLKKNAPLRKYIPBMLQAEITDLRLNNEPRN-----YLCAQE 423  
Db 402 PESGRSM-----LGPQACMHTLPPDTPSGRPL-IENSEGTIDAAASQSFVTPSSSR 452  
Oy 424 AVMLEIQ-----QINDYRSKRTGLGSLYGENDLLGLDGDPL--REROMAEKOLAAL 474  
Db 453 TSNMLNLSFGLHGFKEQSHLKRSSSIDALVADSEGGSGSEPLCYAVAGSOSSPRTGLPS 512  
Oy 475 GDILSKYEEDRSAPMDFAVNTFMS-----H-----AGIRLRESRS-CTAEKTQSPADK 522  
Db 513 GDEIDSEFTNTEPCNISRFSLSLSTLSKESLIGIR--SRYSQSSPKISGSKSR 569  
Oy 523 -----DKMLPFPPTKKQSSNSK----- 540  
Db 570 LVRFDTVCSYSEORRSYFQEPPEGKRIQEEBDEVEYIIPAKSESEKYYKVSRTSFLNRM 629  
Oy 541 -----KEKDALEDKRN-----PLIRYIGKRS--SSQSIKPGVNRNIHQ 579  
Db 630 TSPRNKSKMNKDKTEKEKKNRHHQFVPGTSGVLYQCGCGCKTLLGKRSLOCANCK----- 684  
Oy 580 FENSHQ--YDVEPGTORLS--TGSPFEDLLSDSSRSRSTIRGRS-----GSLKG 625  
Db 685 -ANTHKCKDAVPPCTKKFQEKYKKNKPOSITLSSSVRDYAPAGLSLHPSSMPIGIPAG 743  
Oy 626 REE-----MKRS--RKAENVPRPSDYDMAAAEAALHQA--SSASLSTSRSLR 673  
Db 744 RKEFAAOVHPLRSVPETTLSEFRRAVTSLESKDSWRSRSHSDELFQSMGSSPSTESF- 802

QY 674 NTPTPTTKMGKRSTIESPMLGFCSTVILPLHLEDDLQGLSDLEPRPEVQNMONTGKDV 733  
 Db 803 -----MMEDVDSLTWD-----LSSAQEFEAESWLVVDPSFC 837  
 QY 734 ANLTQREIDROEVINELFVTEASHLTPLVDLILEYQRMKKNLMPREELARFNLPEL 793  
 Db 838 SHQEKDVIKRDIIVFELMQTEVHIIQTLLIMSEVFRKGMKELODHSVDKIFPCLODL 897  
 QY 794 IEIHNSMCEAMKMLEEGP-----ITRDISPMLARPDGARELOQVAAPCSYQSV 846  
 Db 898 LETHNHFFPSMKERQESCSGSDRNFVLMQIDILVQOSEENASMKKITYGEPFSHKE 957  
 QY 847 ALELIRTKORKSRFQLEQAEHSPOCRRLDRLIVSEMRLLTKYPLLENIIRKTEG 906  
 Db 958 AMSLKEKLOQNK-KFONFIKINSNLLARRGICPELITLVQITKPYPLVERIQTYTE 1016  
 QY 907 GTSSEHKLCRADQOREILKLFVNEAVKQTEFNHRLLEGYQKRLDATALAESAPLAEPKS 966  
 Db 1017 RFEHRDLCKALGLIKDMAAAYDLKVEYKKNOKMLETINKTNTYTKLKGHVFRKQA 1076  
 QY 967 IDLTTRKMIHNEGPLTWIRISKDKTLDLVLLLEDLVLLQROEERLLKCHSKTAVGSSDS 1026  
 Db 1077 LLSQERALLHDGLVYWKATATGFRKIDALLLTDLVLLFQEKQKITYF-----AAVDQ 1128  
 QY 1027 KQTESPVKLNAVILRSVATDKRAFFITCTSELGPPQIYELVALTSDDKNIMELLEAV 1086  
 Db 1129 KFS---VISLOKLIREVANEERGMFLISASSAG-PEMEIHTNSKEENMMRRIQAV 1184  
 QY 1087 QNATKHPGAAPRPIRHPSPGSEAPVQGSTSRVEINDESVYHTEKPKLPECGGPBOR 1146  
 Db 1185 ESCPREEG-----GRTSE-----SDERKKAARAKI-----QQ 1214  
 QY 1147 VQDKLLIAGFVQDEDEELTLPRAPPSLDENKRGITRPVLLATLGPLMEGLADA 1206  
 Db 1215 COE-ILSNDDQICTYLEKHL-----IYAEIGEL-----S 1244  
 QY 1207 ALEVDENLRHLILMSILRGHTVKTQAGEREDLTPTPSVSITSHMPQSGPGQAPTI- 1265  
 Db 1245 GFEDVHLERHLII-----KP-----DGGEPDPAASIL 1271  
 QY 1266 -----SDSTRLL-----RPEGSS---QPEGEDVAVSSILANLPPR-TRSS 1299  
 Db 1272 AALAREAESLQVAVKASKMGDVQSSESPGCTVLMDPSTQDVAPASPLAVTEGTEGR 1331  
 QY 1300 GWNDSPELDNRNMAAASSTEPASAKYVAKVSLLEGCGGAKAVAGSNAIPDSQ----- 1354  
 Db 1332 GCWDV-----DGLQGVVTDLAVS-----DAGEKVEYR 1359  
 QY 1355 --SESELSEVEEGGAQTGNCFFVSPAGPLDSS-----TEPTGTPRPSQCHSLPA 1403  
 Db 1360 SFSGSSGSSITLQINLPLRLYSLOALTIQDSHLEHKLVLQGNESLAPS---HSFNG 1415  
 QY 1404 WPTPEQPRRGVGGCCSLVRRDVIVHTITQLTIKILKLMELANHELKLS 1457  
 Db 1416 GPLQDQERSRYLEKQREELA--NIHKLQHQFOEQORFMRHRTCDQOOREQAEQS 1467

RESULT 4  
 I49342  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 20-Jun-2000  
 C:Accession: I49342  
 R:Whitehead, I., Kirk, H., Torgoun, C., Trigo-Gonzalez, G., Kay, R.  
 J. Biol. Chem. 270, 18388-18395, 1995  
 A:Title: Expression cloning of Ifc, a novel oncogene with structural similarities to guanine nucleotide-binding proteins  
 A:Reference number: A57387; MUID:9535462  
 A:Accession: I49342  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-573 <RES>  
 A:Cross-references: EMBL:U028495; NID:9902022; PTDN:AAC52234.1; PTD:9902023  
 A:Genetics:

A;Gene: lfc  
C;Superfamily: pleckstrin repeat homology  
F;472-570/Domain: pleckstrin repeat homology <PLK>

Query Match	5.98; Score 462; DB 2; Length 573;
-------------	------------------------------------

Matches 119; Conservative 113; Mismatches 173; Indels 50; Gaps 11;

QY 659 SASSASSLSTRSLENPTPPFTPKMGRSIESPNLGCTDVI-----LPHLEDDL- 709

```
Db      142 SLSLAKSVSTNIAGHFNDSP-LGLRQILSQS-----TDSLNMNRNRTLSVESLIDEGVE 196
```

0Y 710 ---GOLSDLEP---EVONTWOHTVGKDVANLTOREIDPROEVINELFVTEASHLRTLR 762

Dh 197 VEYNEMSFEMDEKDEADSWSLAVDSSEI.OOHKKKVKKODVYELTOTELHHVBTLK 256

00 763 YDI TEVOPMBKENI MBPEEI ABIEBNI DEI TETIUNSCFEAMKRI BEEG--D-----TT 814

[illegible]

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
84

[illegible]

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

8/3 KRLZKRLI VSEMYKLI KIFLEENI I KHI EGGI SEHEN ECK KANDYCKREI UNF VNEAVK 334

D5 371 KRHGVQECILVYQRIYKPYLNRILQNSHGVEEFYQDGLASALGLVKELLSNVDQVHE 436

QY 935 TENRHRLEGYOKRLDATALEASNPPLAE--FKSLDLTRKMIHEGPLTWIRISKDKTLDL 992

Db 437 LEKEARLQEIYNRM-----PRAQTPVPGKGPFGRDELLRRLIHGCLLWKTATGRFKDV 492

QY 993 QVLLLEDVLLQREERLLKCHSKTAVGSSDSKQTESPVLKLNVLIRSVATDKRAF 105

Db 493 LLLMTDVLVLQEKDQKIETSLDKPS-----VSLQNLIVRDIANQAKGMF 5400

QY 1053 ICTSELGPPQIYELVALTSSDKNIMWELLEAVQ 1087

```
Dp      541 LISS--GPEMEYEVHAAASRDDRTTWIRVIOOSVR 572
```

RESULT	5
T30434	

P47 LBC oncogene - human

```
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Nov-1999
```

R;Toksoz, D.; Williams, D.A.

A:Title: Novel human oncogene 1bc detected by transfection with distinct h

A:Reference number: 138434; MUID:94119604  
A:Accession: T38434

A/Status: preliminary; translated from GB/EMBL/DBJ

```
A;Residues: 1-424 <RES>
```

C;Superfamily: CDC24 homology

2/17 2/14/2011 10:21

Query Match 5.48: Score 428: DB 2: Length 424:

Best Local Similarity 25.7%; Pred. No. 5.be-14;  
Matches 100; Conservations 103; Mismatches 150; Indels 36; Gaps

**THE UNIVERSITY OF CHICAGO**

[illegible]

1  
 2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525

QY	100	110	120	130	140	150	160	170	180	190	200	210	220	230	240	250	260	270	280	290	300	310	320	330	340	350	360	370	380	390	400	410	420	430	440	450	460	470	480	490	500	510	520	530	540	550	560	570	580	590	600	610	620	630	640	650	660	670	680	690	700	710	720	730	740	750	760	770	780	790	800	810	820	830	840	850	860	870	880	890	900	910	920	930	940	950	960	970	980	990	1000
100	110	120	130	140	150	160	170	180	190	200	210	220	230	240	250	260	270	280	290	300	310	320	330	340	350	360	370	380	390	400	410	420	430	440	450	460	470	480	490	500	510	520	530	540	550	560	570	580	590	600	610	620	630	640	650	660	670	680	690	700	710	720	730	740	750	760	770	780	790	800	810	820	830	840	850	860	870	880	890	900	910	920	930	940	950	960	970	980	990	1000	

DB 98 GVISQGMADLLEEQVMEXLEFPCDELISLHSQFEQRLERKESLVDSKNEELAKRI 137



[illegible]

Db	559	CNSSTEGHGL-----	TLGPEKEXKIPKSLSEKVNESKTELENGVPLSE	604
Qy	541	---KEDMALEDOK-----	-NPIIATYIGKFRSSQIKPGVNRNIIQ--HFENSHQYVPP	591
Db	605	ILGKEDRTDEOELMSPGCTIKFRSSIGKESQEVVRRSKGCIENSTWATFEESQHLP	664	
Qy	592	GTORLSTGSPFEDLLESOSRSRIIRLGRSGSLKREEMKRSKAENVPRRSDVMDAAA	651	
Db	665	GAE-----DOMLERLYEKEDSPRRPEEDQACRLOKENEQPIGEYEA-	711	
Qy	652	EAARLHOSASSASTLSRIENPTPTFRKMRGRSIESN---LGF--CTDIYILPHLE	706	
Db	712	--GQILELLEIKESQESLRS-----PEEDQAGRSLQKQNOPLCYEAEOQMLERLIE	764	
Qy	707	DDLGLSLDEPEPEVQNMQHATVGRADVANTLOREIDBOEYINELAFY-TEASHLRTLVLD	765	
Db	765	KE-SQESLSKPEE-----NQRIKGLPERE-NQMSIARLENOQTYFPLBSRMRPLRSIE	817	
Qy	766	LIFYRMKKNEMLRRELARLFPULPILIHNSWCAMAKKLEBEPITRIDISDPLMARE	825	
Db	818	V-----EEERQIRVFKLEYSO-----DSLSTLAEEVY-----	845	
Qy	826	DGPAREELQOVAAPFCSSQSVALLERITXOK-----ESR--FOLFMOEASHPCRLT	877	
Db	846	----QPLRYLEEDCINKS-----LEEDKTHKSGSLIEDNGOSIIIPESETYQSLARP	896	
Qy	878	QLRDLIVSEMRILKYPLLENIITKHTEGOTSEHEKLCRARDOCREILKFVNDAYKOTEN	937	
Db	897	EED-----ORINHLKESQFERSRSEEDQ-----VMERSLEGEN	933	
Qy	938	RHRLEGYOK--RLDATALERASNPILAEFSLDITTRKMYHEOPLWRISOKTLDIOLY	995	
Db	934	HESLTSVEKEDQWMSQLEKESQSG--KSLDEQOFT--GPL-----EKENAESLRL	984	
Qy	996	LEEDLVLLQOEERLLKCHSKTAVGSSDSKOTESPVLKLNVLIRSVATDKRAFFIIC	1055	
Db	985	AGOD-----QEOBKLEOETOOQLTRAVNGEOWAASPEKVPDELPRPLGNQOE-----I	1032	
Qy	1056	TSELGPPQIYELVALVTSDDKNIM-----MELLEENVONMTRKGAAPL-----	1096	
Db	1033	ARSLCKEKOESLVSL--KEKGIETVWSLEPIETLETAEBEDLERKKSIDTOEPLMSTEV	1090	
Qy	1099	-----PIHPSPSQSBPAYOGSTSSRVELINDSEVYHTEKEPKLPE--GP--GREORVODK	1150	
Db	1091	ARETVEPEDEPPSS-----LGS-----VDENRETLISLEKESQELSSLGKMWELTRVEDS	1141	
Qy	1151	QLINQGER--VOEEDDEBELTIPRAPSLDGENK-----GIRTRDP	1188	
Db	1142	QOCQIQVEGLOEBQHOSLEBVAKOELPSSQNOGRMEDVEYGRAGVQOEAFLATTGCTEDK	1201	
Qy	1190	VLLALTGPLLMEGLADALADEVENTLRLI--LWSLLPGHTVTKQAAGEBEDULTP-----	1242	
Db	1202	AELHLRG-----QGEEBEBAABEGELLDIYGEAMSL-----GSSEPKBEYRYAEALD	1248	
Qy	1243	-----TPSVYSITSHPMDDPGSPGQAPITSDSTYL--ARPESSQPEGEDVAVSSLA	1290	
Db	1249	NLEGGALEVPVAQSMPEVTEREDRDRAQAGODSIEVTLGLEAARGLLELQEVVGLIEDPR	1308	
Qy	1291	H-----LPPRTSSGWDSPEDLRNDPAEAAASTEPAAASYKVVRRKXSLIPGGVGAAVA	1344	
Db	1309	HFAREEAIPR---SLGE-ESVAKKINQGLEPKPEK-----KEGALDSGILLELPKTS	1357	
Qy	1345	GSNAIPDSGQSESELSEVEGGAOTGNCYVSNPAGPLDSTEPTGTPPSPQ-----	1397	
Db	1358	-SEALBECOGHDESE--SMEGMBEEE-----ASLETSDHGSDAPQPRPETEEDGQAQA	1409	
Qy	1398	-----CHSLP-----AMTEPOPYRGVY-----GGQCSSILYRVNDVYIHTI	1434	
Db	1410	LTAAGPKLLEPCSTIPLIDAHLEQDPA--BGIOEAGWQDEASEALEREENPEPGIGEI	1468	
Qy	1435	EOLTIKIRHLKDELALHRELKSLIGESSGGTIPVQSFTEAA--RWT---DYSLSP-----	1486	









```
Db 483 DEMESLFSLPMLPEFQKVFLETL-----EDG--ISASSDFNTLEMPQSRKLLFLSGGSF 536
Qy 841 CSY-----QSALELIRTKOKRESFQLFMOEAESHPOCRRLQDLIVSEMQ 888
Db 537 LYVADHFYLSGFCANHKVQKLERAKTDKAKFAL--DARNPTQHSSTLESYLKIPVQ 595
Qy 889 RLTGKPLLENTIKHTEGTSSEHKLCRARDOCREILKFNVAVKOTENRHRLEGYOKRL 948
Db 596 RVLKPLPILKELVSLTDOSEEHYHLETKAMEKVASINEMOKIYED-----YGTVF 649
Qy 949 DATAERASNPILAERFSIDLTRKMHIEGPLTW-----RISKDKTLDLQVLLLEDL 1000
Db 650 DQLVAEOS-----GTEKEVTELSMGEILMHSTVSWLNPFLSLGKARD--LELTVEFKRA 703
Qy 1001 VVLLOROEBRL--LLKCHSKTAVGSSD-----SKOTFSPVLKTNNAVILRSVADKRAFF-- 1052
Db 704 VLLVYKENCCKLKKLPNSRPAHNSTDLDPFKFRMLIPISALQVRLGNPAGTENNSIMWL 763
Qy 1053 IICTSEL--GPOIYELVALTSSDKNIMLEAV-ONATKH-----P 1093
Db 764 IHTKIEIGREPTIFQLCCSDSESKTNIVKIRSLIRENFRHICELPLETKCKDRLVP 823
Qy 1094 GAAPPIHDPSPSQE-----PAYGSTSRVEINDSEVYHTE 1131
Db 824 LKNRPVPSAKLASRSRLKVLKNSSSNEMWTGETGKTLTLDSDGSLSSGTQSSGCPTEGR 883
Qy 1132 KKKPLPEGPEGEORVOD--KQLIAGEPVQOEDEEELTLRPAPSLDGENRGITR-- 1187
Db 884 QDSKTSPOCKYHPGLADPNLIKESDILSDDDHROTVAQGSTKIDEIFQFORLKS 943
Qy 1188 -DP 1189
Db 944 EDP 946
```

```
RESULT 13
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N:Alternate names: hypothetical protein EG:49E4.1
C:Species: Drosophila melanogaster
C:date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C:Accession: T13564
R:Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A:Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A:Reference number: 217689
A:Accession: T13564
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5327 <SPA>
A:Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C:Genetics:
A:Cross-references: FLYBase:FBgn0025392
A:Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A>Note: EG:49E4.1
C:Superfamily: Drosophila 576k microtubule-associated protein homolog
```

```
Query Match 3.2% Score 255.5; DB 2; Length 5327;
Best Local Similarity 19.0%; Pred. No. 0.00076;
Matches 334; Conservative 276; Mismatches 722; Indels 423; Gaps 76;
```

```
Qy 10 DRSAKKOSHLSRP--IASWLSLSLSDSTPERTPSHHROPDTSSETTAGLVORCVII 67
Db 2310 DKSATVTSKSVSRPLSVASDHEAAVAIEDDAKSSISPKKSRGFAEIVSPPIEBA--- 2366
Qy 68 QKDQGFQFTVSGDRVLVQSVRPGGAAMKAGVKEDRIIKNGMVTWTSLSHLEVYKLIK 127
Db 2367 -----TMEFSKIEVEK-----SSLALSLOGS-----GKLDQTSDEPDAVEGDF 2407
Qy 128 SGAVYALTLTG--SSPSPVGVSGLOQNPVAGVLVNPPIPPPPPLPPPOHIT-CP 183
```

```
Db 2408 SHAVASVTVPTTLTKPAELAQIGAAKTVS-----SPLDEALRTP--SAPEHISRAD 2457
Qy 184 KPLQ-----DPEVOKHATQIWL-----NMLROBEELDQILRP----- 216
Db 2458 SPAEACASEIASQDKSPQVLKSSSRPAVYAEKSDAAQLKSSVEDLRSVASTEISRPAS 2517
Qy 217 CGETSORTCE-----GRLSVDSQAD-----SGID 241
Db 2518 AGETASPIIEEAPKDFAEFEQAEKAVLPLTIELKGNLPLTSSPVVAHVASVPAELSKVD 2577
Qy 242 SGTERRPPTSISELMNRNSVLDSPGDSPOTSPVILARVQHHRGSDALLP----- 294
Db 2578 IKTASSPIDEA---PKSLIGSPAEBRPE--SPAESKDAEVEKSKDAPRPSPVYESIK 2633
Qy 295 --LHNGIDQSPKPLIIGPEEDYDPGYFNN-----SDILFQDLERIKSHPAYLVFLRYI 348
Db 2634 ADSTGDISPSPESVLEGGKDDVEKSKESRPSVSASITGDTVDVSPASVSVKDE 2693
Qy 349 LSOADPGPLFLYLCSEVYQOTPKQSRSLG---KDINIFLEKNAPLRVKIPENLQAEI 404
Db 2694 HDKASRRESIAKVESVIDEAGKSDSKSSODQDEKSTLASKESRRESVYESSKDA 2753
Qy 405 DLRLNNEEDPRNVLCQAO-----EAVMLEIOQINDYRSKRL 442
Db 2754 E---KSEGRPESVIVASGEVPRESKSPLDSDKTSRPGSVYESVTAE--DEKSEQSRRES 2808
Qy 443 GIGSLYGE-----NDLGLDGDPLRERQM---AEKQIALAGDILSKYE 482
Db 2809 VAEVSKADTKKQKQSGEASRPSSVDELLKDDDEKQSRQISITGSHKAMSTGMD----- 2862
Qy 483 EDRSAPMDPAVNTFMNHAGIRLRRESRSSCTAEKTSQAPKDKMLPPFPYTKR-----QSS 537
Db 2863 ---ESPMDKADKS-----KEPSRPEVASIEIHENTKDESPGSRDSDVABISIKSD 2911
Qy 538 NSKKEKDALEDKRNPIILRYIGKPK-----SSSQSIGKNVANIQHRENSHOYDV 588
Db 2912 ITKGEKSPLEKSEVSPESVYSIDKEKASRESVAESVKP-----ESSKDANS 2961
Qy 589 PEPGTQRLSTGSFPEDDL-----ESDSSRSSEIRLHGSLGKREMKRKRAENPPRPS 643
Db 2962 APPSKKH-----SRPESVLSLKDDEGDKTTSR--RVSVADSIKDEKSLVYQGEAS--RPES 3013
Qy 644 DVD--MDAAAEARLHQSSASSASSLSTRLENPPPPPKMGRRRESPNLGFCTDVL 701
Db 3014 EHESLKDAHA-----PSQETSRPESVYTESVADGKSPVASKASRPAS-----V 3056
Qy 702 PHLEDDQLQSLDLEBEPEVQNMQHTGVQDVANILTQREIDROEVINELFTVASHLRTL 761
Db 3057 AENAKDSADESKQRPESLPQSKAGSI-KDEKSPLASKD-----EA----- 3096
Qy 762 RVLDLITVQRMKKNIMPEELARLFPNLPLELITHNSCEAMK-----KLREGPII-RD 816
Db 3097 -----EKSKEES--REESVAEQFPVLVSKESVRPASVAESVKQDAEKSKEESPLMSKE 3146
Qy 817 ISDPMLARFDGPARELQO-----VAQCFQSYQVLALELR--TKORKE 858
Db 3147 ASRP--ASVAGSVKDAEKSKEESRESVAEKSPLPSKASLRPASVAESVKQDAEKSKEE 3204
Qy 859 SRFOLEMOEAESHPOCRRLQDLIVSEMQ-----LTRYPLLENITKH 903
Db 3205 SRRE--SGAEKSPILASKASRPASVAESIKDAEKSKEESRESVAEKSPLPSKEASRP 3261
Qy 904 TGGCGTS-----EHEKLCRARDOCRE-----ILKFVNEAVKOTENRHRLEY 944
Db 3262 TSVASVYKDAEKSKEESRDSVAEKSPLASKASRPASVAVASVQDAEKSKEESRESV 3321
Qy 945 OKRLDATALERASNPILA-----AEFSLDLTRKMHIE--GPLTWIRISKTDLDLOVL 995
Db 3322 AEK-SPLATKEASRPASVAESIKDEAE-KSKEESRRESVAEKSPLA--SKASRPTSV- 3375
Qy 996 LLEDVLVLLQROEBRLLLKCHSKTAVGSSDSKOTFSPVLKTNNAVILRSVATDKRAFFIIC 1055
Db 3376 -AESVKADEAKSKEE-----SSRDVSAEKSPLASKASRPASVAESVQDAEKS 3423
```

QY 1056 TSELPPQIYELVALTSSDKNIMMELLEAVONA--TKHGGAAPLPIHPSPGSEPAYQ 1113  
 Db 3424 KEERRESVAKSPPLASKEASRPASVAESVVDKAEKSEERRESVAEKSPPLASKEASRP 3483  
 QY 1114 GSTSRVEINDEYVHTKEPKLPE--GCGPEQVQDKQLIAGEPVOEDE--EELRT 1169  
 Db 3484 ASVSESVK--DEAEKSEKESRESVAEKSPPLPSKEASRPTSYA--ESVKDEAEKSEKESRR 3540  
 QY 1170 LPRAPSLDGNRGIPTRDVPLATLPTLMLEGLADALEVENLRHLILMSLLPGHTVK 1229  
 Db 3541 ESAVAKSSPLASKEASR--PASYA-----ESVKDEAEKSEKESRR-----BSVA 3581  
 QY 1230 TOAAGEBEDLPTPTSVYSTSHMPDPSGQAPATISDTRILARPESQGE--GEDYAVS 1287  
 Db 3582 EKSPPLASKEASRPASVAESVVDKAEKSEKESRESVAEKSPPLPSKEASRPTSYAESVKDE 3641  
 QY 1288 SLAHLPTPTSSGVWDSP----ELDR--NPAAEAASSTPPAASYKVVYSLPLGGGVCAAK 1342  
 Db 3642 ADSKEESRESRESVAEKSPPLASMEASRPTSYAESVKDETEKSEKESRESYTERKSPPLPSKE 3701  
 QY 1343 VAGSNAIPDGSSESELSEVEYEGAGATGNCIFY-----SMPAGPLDSTPTGTPTPSQ 1397  
 Db 3702 ASRPTSYAESVVDKAEKSEKESRESVAEKSPPLASKESSRPASVAESTIKDEAGTKEOESR 3761  
 QY 1398 CHSLPAMPTEPPQYRGVGGQCSLVARDY--DVIFHTTEOLITKLH--RLKDMELAHR 1452  
 Db 3762 RESMP-----ESGKAESIKGQ--SLASKETSRPDSVYESVKDETEKESGSAIDKQVASR 3816  
 QY 1453 ELKSLGSESSGTTPTVGSHTTEAKRTDYSLSPPAKALASDSQNGEDGSCP--EFG-- 1509  
 Db 3817 P--ESVAVSAKDEKSPHLSREPSYA-----DKSPDASKE--ASRSLSAVETASSPIEGPR 3868  
 QY 1510 --SDIALEDSATDPA 1522  
 Db 3869 SIADLSLPLNLGTGA 3883  
 RESULT 14  
 S21424  
 nestin - human  
 C:Species: Homo sapiens (man)  
 C>Date: 20-Feb-1995 #sequence\_revision 20-Feb-1995 #text\_change 05-Nov-1999  
 C:Accession: S21424  
 R:Author: J. McKay, R.D.G.; Zimmerman, L.B.; Lendahl, U.  
 Submitted to: the EMBL Data Library, May 1992  
 A:Description: Characterization of the human nestin gene reveals a close evolutionary re  
 A:Reference number: S21424  
 A:Accession: S21424  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1618 <DAH>  
 A:Cross-references: EMBL:X65964; NID:g35018; PIDN:CAA46780.1; PID:g35019  
 Query Match 3.1%; Score 247.5; DB 2; Length 1618;  
 Best Local Similarity 20.1%; Pred. No. 0.00033;  
 Matches 319; Conservative 192; Mismatches 539; Indels 541; Gaps 76;

Db 195 RVAHMETSLDQREBRLARAVOGAREVRLQLQAEKSGLLERRALAEQRLBGRQERLR 254  
 QY 410 NNE-----DPRNVLCAGAEVAMLEIOEQINDIRSKRTLGGS----LYENDILGL 456  
 Db 255 ATEKPOLAVALOEKQOGKOGISOIAQVLEGKQOALHKMSLSLEVAYVRTLLEAENSRLOT 314  
 QY 457 DG-----PPLREQMAE-----KOLALDILS--KYEDRSAPMDFAVNTSMHA 500  
 Db 315 PGGSKTSLSFQPKLELOPPTPEGRRLGSLPVLSPSLPPLATLTTPYPALILKNQ 374  
 QY 501 GIRLESRSSCTAE--KTQAP--DKDWLPFFPKTKKOSMSNKE--KDALDKRR 551  
 Db 375 --EFLQARTPTLASTPIPTPPQAPSPAVDAEIRAQDAPLSLITOGGRKQAPPLRAEAR 432  
 QY 552 NPLIRYIGRKSSSQSRIKPNVNIIOHFNSHOYDVEPGTOR--LSTGSFED----- 604  
 Db 433 VAL-----PASYLPG-----PEEGGGRQEAQSTQSPEDHNSLA 466  
 QY 605 --LLESDSRSEIRIGRSGSLK-----GREEKRRKAKENYPRPSDY---- 645  
 Db 467 PPLSPDHSLEAKDGSGSRVPSICRGEQGIWGLVEKETALIEKVYSSLOQELIMEE 526  
 QY 646 -----DMDAAAEARLHQASASSASSLSLSTLENPTPTPKGGRSIESPNIQFCT 697  
 Db 527 DLNRKEIQQVPLEKETTSLGEEIOESLKT--LEN-----QSHETLERENQ-- 573  
 QY 698 DVLPHLEDDQLQSLDEPEPEVQWQHVGDVYANLQREIDQEVINELFYEAASH 757  
 Db 574 ----CPNSLEEDLTLSLEKE-----NRAIKGCGSSTSKRCGRQIKPTGKED 620  
 QY 758 LRTLRVLDLIFYORMKEN-----LMPREELARLPMLPELIEFHNSMCEA 803  
 Db 621 TQTL-----QSLQKENELKMSLEGNETLFLFTEGNOELVSSLOENLES----LTA 668  
 QY 804 MKRLREBGPITRIDISPMIARFDPARELOQVAAOFCSYVALBELIRT--KORRESR 861  
 Db 669 LEKENOEPLRSPVGDDEALR--PLTKENQ-----BPLSLDENKEARR 711  
 QY 862 QLFMOEASHPOCRRLOLRDLIYSEMOURLKYLLNITIKHTEGTSHEKICRARDOQ 921  
 Db 712 SL--EKENOEPLKTLDEED-----QSTVRLPE--TENKSLRSLEED 750  
 QY 922 RELIKFVNEAVKOTENRHRLEGYOKRLDATALERAS--NPLAEFKSLDITTRKMIHEGFL 980  
 Db 751 QETLRTLE--KTQQRRSLSGQDMTLRPPKVDLEPL--KSLD----- 791  
 QY 981 TWKISDKTLDLQVLLLEDLVLLQOEERLL--LKCHSKTAVAGSSDSQOTSPVLKLAN 1038  
 Db 792 -----QETARPLENENQEFKSLKEESVAVKSLETE----- 823  
 QY 1039 VLIRSVATDKRAFFICTSELGPPQIYELVALTSSDKNIMMELLEAVONATKHPGAAPI 1098  
 Db 824 -----ILLESLSKAGQENLETKSPETQAPLMTPE 852  
 QY 1099 PIHPSPGSEPAYQGST-----SRVEIN-----D 1124  
 Db 853 EIKKS--GGNESSRKNSRTITGVCSEPRDIQTPGSGEIGIIFISSMPEGETISRGVD 910  
 QY 1125 SEVYHTEKEPKKLPEBGPGE--QRYQDKQLIAGEPVOE--ED--BEELRTPLRAPSLD 1178  
 Db 911 KESQRMLEEEENIKGEYSELSLSEEGQELPQSDADVQRMEDYVTKDDELQAESPPGMA 970  
 QY 1179 G--ENR-----GIFTRDVPVLLATLG--PLMEGLADALADLV-----ENLRHLILMSL 1222  
 Db 971 GYENKDEALNLEEDP-----GFTGKEEVYEQGLNAT--BEVMPGCGHPEDEPKQRLG 1025  
 QY 1223 LRGHTVKTQAG-----BP--EDDLTPPTSVYSTSHMP 1254  
 Db 1026 VEGASVYKAGABGLQDPREGSQVGTPTGLQAPGGLPAETPLVEDVAP-----CG 1075  
 QY 1255 DPESP-----GQAPITSDSTRILARPBSQPEGEDVAVVSLAHLPPTRSSGWDSPSLDRN 1310  
 Db 1076 DQASPEVMLGSEPAKMGESAAGAEPLGGGVG-----GLDPPGHLTRE 1117







**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: September 18, 2002, 10:35:58 ; Search time 20.02 Seconds  
(without alignments)  
1863.031 Million cell updates/sec

Title: US-09-695-795-4  
Perfect score: 7883  
Sequence: 1 MSIRLPHSIDRSASRKQSHL.....EGSDIALEDSATDFAVPCP 1527

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: Issued\_Patents\_AA:\*  
2: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1506.5	19.1	1461	2	US-08-993-228-10
2	1196.5	15.2	912	4	US-08-943-768-2
3	318	4.0	1805	1	US-07-853-913-2
4	258.5	3.3	710	4	US-09-079-812E-2
5	247.5	3.1	1618	1	US-07-853-913-4
6	239	3.0	2101	1	US-08-466-390-4
7	239	3.0	2101	1	US-08-470-850-4
8	239	3.0	2101	1	US-08-467-781-4
9	239	3.0	2101	1	US-08-195-487-4
10	239	3.0	2101	2	US-08-483-924-4
11	239	3.0	2101	5	US-09-452-294-1
12	239	3.0	2101	5	PCT-US93-06160-4
13	237	3.0	626	4	US-09-155-770-7
14	203.5	2.6	1780	1	US-08-769-309A-5
15	203.5	2.6	1780	3	US-08-994-570-5
16	200.5	2.5	1346	2	US-08-635-121-2
17	189.5	2.4	2860	2	US-08-826-267-2
18	188.5	2.4	1104	4	US-08-923-992A-6
19	187.5	2.4	2482	1	US-08-328-254-6
20	187.5	2.4	3248	1	US-08-353-700-1
21	187.5	2.4	3248	5	PCT-US95-16216-1
22	184.5	2.3	1164	4	US-08-923-992A-10
23	184.5	2.3	2466	3	US-09-080-855-12
24	184.5	2.3	2466	5	PCT-US94-09943-2
25	183	2.3	2485	4	US-09-290-640-46
26	179.5	2.3	1886	4	US-08-938-105-3
27	179.5	2.3	2843	1	US-07-741-940-2

28	179.5	2.3	2843	1	US-08-289-548A-2	Sequence 2, Appl
29	179.5	2.3	2843	1	US-08-452-654-2	Sequence 2, Appl
30	179.5	2.3	2843	2	US-08-370-235A-2	Sequence 2, Appl
31	178.5	2.3	1164	4	US-08-923-992A-2	Sequence 2, Appl
32	178	2.3	1939	4	US-09-310-187A-1	Sequence 1, Appl
33	178	2.3	2465	2	US-08-596-291-3	Sequence 3, Appl
34	178	2.3	2465	3	US-09-100-804-3	Sequence 3, Appl
35	177.5	2.3	2842	1	US-07-741-940-7	Sequence 7, Appl
36	177.5	2.3	2842	1	US-08-289-548A-7	Sequence 7, Appl
37	177.5	2.3	2842	1	US-08-452-654-7	Sequence 7, Appl
38	177.5	2.3	2843	1	US-08-452-655B-2	Sequence 2, Appl
39	177.5	2.3	2843	1	US-08-452-655B-7	Sequence 2, Appl
40	177.5	2.3	2843	3	US-08-450-582-2	Sequence 7, Appl
41	177.5	2.3	2843	3	US-08-450-582-7	Sequence 7, Appl
42	177.5	2.3	2973	2	US-08-821-355A-7	Sequence 7, Appl
43	177.5	2.3	2973	2	US-09-003-687A-7	Sequence 7, Appl
44	177.5	2.3	2973	4	US-09-136-605-7	Sequence 7, Appl
45	177	2.2	984	1	US-08-242-932-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1  
US-08-993-228-10  
; Sequence 10, Application US/08993228  
; Patent No. 5976838  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallee, Edward R.  
; APPLICANT: Racie, Lisa A.  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Agostino, Michael J.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: MA  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/993,228  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sprunger, Suzanne A.  
; REGISTRATION NUMBER: 41,323  
; TELEPHONE: (617) 498-8284  
; TELEFAX: (617) 876-5851  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1461 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-993-228-10

Query Match 19.1%; Score 1506.5; DB 2; Length 1461;  
Best Local Similarity 29.6%; Pred. No. 3e-105;

Matches 459; Conservative 243; Mismatches 494; Indels 357; Gaps 54;

```

QY 161 VNPILPPPPPPPLPPPOHITGPKPL--QDEYQKNAQTQILMMNLROEEELODILPPCG 218
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MSPILSPHSPGASGNMERITSPVLGEEENNVHNOVELRKMLOKEGRLOL-----Q 56
QY 219 ETQQRCEGLSLDSQPADSGLDGTERF-----PS-----ISELMN 256
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 EDYRRPRAQRLKEIOEAKKHIPLOEOLSKATGSADGAVVTPSRPLGDTLVSESETD 116
QY 257 RNSVL-----SDPGLDSPQTSPIILARVAQHRRQSDALLPLNHQIDQS 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 PGVLGDTGSSGDASPSDDNADSPKSGP--KERITLSENPKESE--TIDDTQTSLVGS 173
QY 304 PK---PLITGPEEDYDPCGFNNESDI-----FQDLKLSHPAYLVFLRTLSQAD 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 PSTRIAPHITGAEDD--DFGTEHEQINQCCSFQISLTKSPAHILAVLHHVVSQFD 229
QY 354 PGFLFLFLGSEVYQOTNPKRSISGKDINWIFLEKNAPLRVKIPEMLQAEIDLR---LRN 410
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 PATLLCTLVSLDKYKHTSKETRRIFLEHOFILNRSANLVSVDEKSADELRKRPELIP 289
QY 411 NEDPRNVLCBAQVAMLEIOEQINDYRSKRTLGSLYGENDLGLDGPLREROMAERQ 470
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 290 EDLHRHYIQMQRVHPEVQHRHLKDFRQKSMGL--TLASELTKLDAERDKRLTIKE 347
QY 471 LAALGDLISKYF-----EDRSAPMDFAVNTFMSHAGIRLRESRSCIAETQOSAP 520
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 348 RTCAEQIVAKIEEVLMXAQAVXEDKSTMQVILWMLPKHLGVVYKEPR-----NLE 398
QY 521 DKCKMLPEFPPTKKQSSNKKEDALDEKRNPLIRIKGPKSSOS-----567
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 399 HKKGRIGFLPKIKO---SMKKDKEGEEKKRRGFPSTILGPKRRPSRIDNSAIGRAMELQK 455
QY 568 -----IKGNVNRNIIQHFENSHQYDVPEPTQRLSTGSPEDLDESRSREIRLGRS 620
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 456 ARHPKHLSPSSVSPPEODSAKLQSGLANEGT---DAGYLPA-----NSMSVASGAS 506
QY 621 GSIKGRBEMKRSKAENVRPRSDVMDMAAEARLHQSSASSLSITSLSENPPPF-- 679
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 507 FSGEGSKE-----NDTGSKOYEVSAPGDTLDGTPRLNT--VFDFPPPLD 551
QY 680 -----TPKGRRSIESPNLGFCTDYLPHLLDDEDLGQSLDEPEPEVON 723
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 552 QVQEECEVRYVTEHTPRPKF--FDSVAFGESQ-----EDQGF--NDLETD--PN 600
QY 724 MQHTVQKDVANLTQREIDRQEVINELFVTEASHLRTLRVLDLIFYQRMKRNMLPREEL 783
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 601 WOOLVSREVLLGLKPCQIKRQEVINELFYTERAHVRLTKVLDQVYQVRSREGIILSPSEL 660
QY 784 ARLEFNPULLEIETHNSMCEMMKRLRE--EGPIIRDISDPLARFDGPARELOQVAAQRC 841
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 661 RKLFSLMLELILQHLINQEMKAVKRKNETSVIIOIGEDLLTFSGPGEKELHAAATGC 720
QY 842 SYQSVALELIRTKOKRESRQOLMOEAESHPOCRLOLRDLIASEMORLTKYVLLLENT 901
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 721 SNQPFLEBMKSKQKKDSRFQTVQDAESNPRLCRRLQDKDIPTQMRRLTKYVLLDNTA 780
QY 902 KTEGSGTSEHEKLCARDOCREILKEFVNEAVKQTENRHRLLEGYQKRLIDATALERASNP 961
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 781 KTEWMT--BREKVKKAADHCROITLVYNOAVKAEKQRLVEDYQRLDLSLKLSTSPVNV 839
QY 962 AEFKSLDTLTRKMHTEGRLTWIRISKQTLDOVLLLEDLVLLLOREBERLLKCHSKTAV 1021
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 840 EELRNLDLTKRKRIHGGPLWAKVNRDKTIDLYLLLEDLIVLLQKODDRILVJACHSKIIA 899
QY 1022 GSSDSKQFSPVLKLAVALIRSVATDKRAFFITCSELGPQIYELVALTSSQKNIMMEL 1081
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 900 STRADSKHTSPVILKSTLVLRQVATDKKALFLVISMDSNG--AQYELVLAQTVSEKTYWQDI 958
QY 1082 LBEAVONATKHPQAPRPIIHPSPPG-----1106
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 959 ICR--MAASVKBQSTKIPILPQSTPGEGDNDDEDPSTLKEBQHSIVTGLQSPDRDLGLS 1017

```

```

QY 1107 -----SQEPAYQGSTSSKVEIN-----SEVYHTEKEPKKLPEG-----PGPE 1144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1018 TLISKRNQSHSLSTSGSEVRLDLFAERQFAKQHTDGTLKEVEGDYQIAIPDSHLPVSE 1077
QY 1145 QR--VOD-----KQIADGEPVQED--BEELRTLPRAAPSLDGEKRGIRTDPPVLLAL 1194
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1078 ERNALDALRNILGLIKOLLVQQLGTLTKESVQEDMWHPR-----YRT-----AS 1120
QY 1195 TGPLMEGLDALEDEVENLRHLILWSLPGHTVKTQOAGEPEDDLPYPSVITSHPW 1254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1121 QGF-----QFDSVYIQNSENK---AYHSGEHNPFRTGTDIATCVSPRSTESFA--PR 1170
QY 1255 DPQSPGAPRTISDSTR-----APRGS--QPEGDVAVSSIAHLPPR--TNS 1298
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1171 D--SVGLAPQDSQASNLVMDHIMTPPEMTPEBGLDSDGSHFPDARAHADENPSEQ 1228
QY 1299 SGVWDSPELDNR-----PAAEAATERPAAKYKVRKYSILPGGVGAUKYA 1344
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1229 DGAVNKEEKDVNLRISGNVILLDGYDPVQDSSTDEEYAS-----SLTQPMTCIPAVEST 1283
QY 1345 GSNA--IPDSGQSESELS-----EVEGGAQATGNCFFVSMRAGPLDSSTE----- 1387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1284 HQQSHSPQNTHSDGALSPFPRFLVQQRWGMAMY--SCFEIQSPSCADSQSOLMEYTHKI 1342
QY 1388 -----PRTGPPSPQCHSLPAMTPPEOPYGVGGCCSLVRKRDVY 1429
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1343 EADLEHLKEGGKLTPEFFAKGMDQPSOT-----STQIKVRAACPGGC-----RLDIL 1391
QY 1430 IFHTI-----EOLTIKLHRLKDMELAHRELKLSGSGSGGTTPVGS 1471
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1392 EYRPCLTTSWLQCCGCHESVT-----EDLPVNHILGLVKSQAGSQTSSXSCGS 1438

RESULT 2
US-08-943-768-2
; Sequence 2, Application US/08943768C
; Patent No. 6238881
; GENERAL INFORMATION:
; APPLICANT: Hart, Matthew J.
; TITLE OF INVENTION: No. 6238881el Nucleic Acids and Polypeptides Related to a
; FILE REFERENCE: 1023-US
; CURRENT APPLICATION NUMBER: US/08/943,768C
; EARLIER FILING DATE: 1997-10-06
; EARLIER APPLICATION NUMBER: 60/029,979
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 2
; LENGTH: 912
; TYPE: PRT
; ORGANISM: Human p115 GEF-Rho
US-08-943-768-2

Query Match 15.2%; Score 1196.5; DB 4; Length 912;
Best Local Similarity 34.2%; Pred. No. 4,6e-82;
Matches 335; Conservative 167; Mismatches 324; Indels 153; Gaps 32;
QY 308 IIGPE--EDYDPCGFNN--ESDIIFQDLKLSHPAYLVFLRTLSQADPGPLFLGSE 364
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 23 IIGAEDEDEFENELETNSEEONSFQSELEVQKRRPAHMLLQHVLAQFPGPLCCLHND 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 365 VVQQTNPKRSRSGKDINWIFLEKNAPLRVKIPEMLQAEID--LRTRNNEDPRLVLCBA 421
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 83 MUGSLGPKRAKKAFLDFHSFLEKTAVALVPVPPNVAFLDRDLRDLISEDVQRRVGV 142
QY 422 QEAVMLEIOEQINDYRSKRTLGSLYGENDLGLD-----GDLPERQMAEKQLAA 473
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 143 VOSQVAVAGROLEDFRSKRLMGKTP--WEQELQALEAWVGRDRASEAREHVAERILMH 200
QY 474 LGDI--LSKYEEDRSAPMDFAVNTFMSHAGIRLRESRSCIAETQOSARDKQKWLPEF-- 529

```

```

Db 201 LEMOHTISTDEKSAAYNAIGLYMRHLGYRTKSG-----DKSGHNFPRK 247
Qy 530 -----PKTKOSSNKKEDALEDKKRNPIKRTGKPKSSOSIKGNVNNIIQ 578
Db 248 KYWGNRRSDPPKTKGLSSI---LDAARMNGEPQVDFRHLKAENVAEKKG----- 297
Qy 579 HFENSHQDVPEPGTQRLSTGSPFEDDLESOSRSEIRLGRSGSLKGREMKRSKAEV 638
Db 298 -----ATDRKGGMV-----SRD-----RNTGAPGDTPEVSLHPLSL 331
Qy 639 PPRSDVMDAAEAARLHOSASSASSLSTRLENPPPTPKMGRSIESPNLGFCTD 698
Db 332 DSPDEPADAPLEL-----GDSSPQPMLESIALPESTDGAER-ESPERG----- 378
Qy 699 VILPHLEDDGQ--LSOLEPER-EVQWQHTVQKDVANLTQRLIDQEVNLEFVTEA 755
Db 379 -----DEGEPRSGLEPERPPGMRELVPDTHSLPKSOVKROEVISELLVTEA 429
Qy 756 SHLRTLRVLDLIFYOMRKRENIMPREELARLPNLPETLIEHNSWCE-AMKKLREGPFI 814
Db 430 AHRRLRLVHLDFPQPMACEFFPLELONIFPSLDELLEVLSLDRMLMKRQSGYLI 489
Qy 815 RDISPMILARFDPARBELOVAAOFCSTQVALILRTKOKESRFOLPMOAEASHQC 874
Db 490 BEIGVLLARFPGAEGSWFOKISSRFCSROSPALFOLAKORKDRFCFAFVQEAESRPRC 549
Qy 875 RLQRLDILVSEMRULTYRPLLENIKTKEGTSHEHELCARQCEILLKFEVNAVKQ 934
Db 550 RLQRLKDMIPTEMORLTKRPLLOSIGQNTTEPT-EREVELAACCRIELHNVQAVRD 608
Qy 935 TENRRLEGYOKRLDATALEASNPAAEFKSLDTTRKMIHEGFLTRISKDKTLDOV 994
Db 609 MEDLRLKDYQKRLDLSHLRQSSDPMLSFKMLDITKKKLIVHEGFLTRVTKDAVEYHV 668
Qy 995 LILLEDVLLQOEERLLIKCHSKTAVGSOSKQTFSPVYLKINAVLINSVATDKRAFTI 1054
Db 669 LILDDLILLORODERLLIKSHSRFLTPDGTMLRPLRTLSAMTEVATDHAFAFVL 728
Qy 1055 CSEELGPQIVELVALTSSDKINIMELLEAVON-ATKHGGAAPRPHSPGSEBPVQ 1113
Db 729 FTWD-QEAOITELVAQTVSEKNNMCALITETAGSLKVPAPSRKP-RPRPSTREPLIS 786
Qy 1114 GSTSRVEINSEVYHTEKEPKLPEG-----PGPEORVOD-----KOLIAOGEP 1158
Db 787 SSENEN---GGRETSPARARTERILISDLRPFGRPEGLATATARKYLSLKOLF---P 840
Qy 1159 VOEEDEEELRPLRAPPSLIDGENRGITRDPVYLLALTGELMEGLAD-----ALEDVE- 1212
Db 841 AEEDN-----GAGPPRPGD-GVPGGGLSPARTQE-IDENLISLEETWKOLELEE 889
Qy 1213 ---NLRLILMILMSPGHV 1228
Db 890 EFCRLRPL--SOLGNSV 906

```

RESULT 3  
US-07-853-913-2

Sequence 2, Application US/07853913

Patent No. 5338839

GENERAL INFORMATION:

APPLICANT: McKay, Ronald D.G.

APPLICANT: Lendahl, Urban

TITLE OF INVENTION: Nestin Expression As An Indicator of

TITLE OF INVENTION: Neuroepithelial Tumors

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Millitia Drive

CITY: Lexington

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02173

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/853,913
FILING DATE: 19920319
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/660,412
FILING DATE: 22-FEB-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603,803
FILING DATE: 25-OCT-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/201,762
FILING DATE: 02-JUN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/180,548
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MIT-4641AAAA
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-9540
TELEFAX: 617-861-6240
SEQUENCE CHARACTERISTICS:
SEQUENCE FOR SEQ ID NO: 2:
LENGTH: 1805 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-853-913-2

```

Query Match 4.0%; Score 318; DB 1; Length 1805;  
Best Local Similarity 20.0%; Pred. No. 5,8e-15;  
Matches 323; Conservative 229; Mismatches 605; Indels 456; Gaps 81;

```

Qy 170 PPPPLPPPHQHTGPKPLDDPEYQKHATQI--LW-NMLROEBEELDILPPCGETSQRTIC- 225
Db 159 PRRPAPPHRIPGPA-----PEVEDLARLIGEWRAGVADYQERVAMHMSSTLGOAERLSQ 214
Qy 226 -----EGRLSVDSQEAQSGDSTGERPPSISESLMNNNSVLSQDGLSPQ-TSPVILAR 278
Db 215 AVRGARECRLEVOQLQADR--DSLQERKALEQRLGR-----WQRLQATDKFQLAV 265
Qy 279 VAQHHRQGSDAALLPLNHQG-----IDQSPKPLIIGPEED--YDPGYFNNESD 325
Db 266 EALQEQKQLOQSOIQLIEGGQQLAHLKMSLSLEVAATYRTLLAENSRLQTPGR-GSQAS 324
Qy 326 IIFQLEKLKSHPALVY---FLKYLISQADP-----GRLFLFCSEVYQOT 369
Db 325 LGFLD-PLK--PNFLGIPEDQYIGSVLPALSPISFSPPLPNTLETPTATKQLQEFQDA 381
Qy 370 N-----PKDSRSLSKDIWNIFLEKNAP-----LRVKIP--EMLQAEID- 405
Db 362 RPTPLASTPIPPISAPCPPAEVAQAQEVPLSLQTOQAPEPLMLKATVPSSAISILPELEE 441
Qy 406 -----LRLNNEDEPRNVLCQAQ-----EAVMLEIQDINDY 436
Db 442 PGKQOQHFPDDLTSLATNLNHNHTLBAKDOESSESVSSIFQDEQOIMELVEKADI 501
Qy 437 RSKRTLGSLXGENDLGLDDPLARENQ--MAEQQLAALD--ILS-KYEEDSAPMDF 491
Db 502 EYKVENSSAOKTQES---GLDTEEPQDSQGLQKTTLALQEBEPLMSLKQNTYTAGEN 558
Qy 492 AVNTFMSAGIRLRSSRSCAETQSAAPDKDKMLPFPKTKKQSSNK----- 540
Db 559 CNSSTEGHLG-----TLEGPEKEKQIPLKSLKEKNVESKTLNENGVYULSE 604
Qy 541 ---KEKDALEDKK---RNPLRLYIGKPKSSOSIKPGNVNIIQ--HFENSHQDVPEP 591

```

Db 605 LKCKEDRTDQELMSKGLTKRFSGLKESQEVVPRSKGKLNESMTAFKEESQHPGLGP 664  
 QY 592 GTORLSTGSPEDLLSDSSRSERILGRSGSLKGREMKSRKKAENPRRSDVMDMAA 651  
 Db 665 GAE-----DQMLELVKEKDOSPRSPREEQACRPLQKKNOPPLGEEA----- 711  
 QY 652 EAARLHQSSASSSLSTRSLENTPTPTKMGRSIESPN--LGF--CTDVILPHLE 706  
 Db 712 --QOILERLIEKSQESLRS-----PEEDQAGRSIQKNEQEPPLGEEAEDQOELRLIE 764  
 QY 707 DDLGQSLDLEPPEPVOMTKGVANLQREIVINELFV--TEASHLRLTVLD 765  
 Db 765 KE--SQESLKSPEE---NORIGKPLERE--NOKSLARLENOQEFVPLESRNORPLKSL 817  
 QY 766 LIFQWRKKNLMPRELARLPMLPELIEIHNWCAMKRLREGPILINDISPMLARF 825  
 Db 818 V-----EEEOQIVKPLEVSO-----DSLGSLEENV----- 845  
 QY 826 DGPAREELQOVAOFCYSOVALELITKQK-----ESR--FQLEMOAESHPQCRRL 877  
 Db 846 ----OPLRYLEDDCINKS-----LLEDKTHKSLGSLERNQDSIIPQSETOVSLRP 896  
 QY 878 QLDLIVSEMOQLTKYLLLENITKHEGTSHEKICLRARDQCRELKVNEAVKOTEN 937  
 Db 897 EEDD-----QRTVNHLEKESQEFSSRSSEEO-----VMERSLEGEN 933  
 QY 938 RHRLEGYOK--RLDATALERASNPAAEFKSLDLTKRMHIEGRLPWIRISKDITLDQVL 995  
 Db 934 HESLSVEKEDQVSEGLEKESQDSG--KSLDESOETF--GPL-----EKNAESIRSL 984  
 QY 996 LLEDVLVLOROEBERLLKCHSTAVGSSDSKQTFSPVLKLNVLNSVATDKRAFTIC 1055  
 Db 985 AGOD-----OEOKEQLEQTOOTLRAVGNQMAVSPPEVDELPKPLNDDE-----I 1032  
 QY 1056 TSELGPQOIEVALTSNDNIM-----MELLEAVQNAKHGGAAP----- 1098  
 Db 1033 ARSLKGNOSLVS--KEGCIETVKSLETFIEETAEEDLERKSIDTOERPLMSTEV 1090  
 QY 1099 -----PIHSPPGSOEPAYOGSTSRVETINDESVYTHEKEPKKLP--GP--GPEORVODK 1150  
 Db 1091 ARETVPEDEPPGS-----LGS-----VDENRFTLITSLKESQELSLGKMNVEVTRVEDS 1141  
 QY 1151 QILAQGP--VOEBDEELTLTPRAPSLDGENR-----GIRTRDP 1189  
 Db 1142 QOOLVQEEGLQEOHOHOLEVQKQELPSSGNQOQWEDVEGKAVGOBAPLATTVGYTEDK 1201  
 QY 1190 VLLALGRLMEGLADALDVENLRHLI--LMSILBGHVTKQAAGEPDDLTLP----- 1242  
 Db 1202 AELHLRG---QGGEEBAALAEGLLDIDIVEANSL-----GSEEPKEQVPAEALD 1248  
 QY 1243 -----TPSVVITSHPMDPGSPGOAPTISDSTR--ARPEGSQPGEDEVAVSSLA 1290  
 Db 1249 NLBGALFVVAQSMPEVTRERDRDAQAGQDSIEVTLGLEAARTGLELQOEVVGLDEPR 1308  
 QY 1291 H-----LPRRTSSGVWDSPELDRNPAEAATERAASKYVRAKYSLLFGCGVGAUKA 1344  
 Db 1309 HFAREEAIRP---SLGE--ESVYKAKIAQGLGEPKEP-----KEAGALDSGLILELPKTS 1357  
 QY 1345 GSNALIPDSGSESELSVEGGAQATGCVFVSNPAGPLDSTPTPTGPPSPSO----- 1397  
 Db 1358 -SEALBEGQHESE--SMGEWEEBE-----ASLETSDHESGDARQPRPPTTEDEGAQAA 1409  
 QY 1398 -----CHSLP-----AMPTEDQRYRGVR-----GGQCSLSLRVDVIVFTTI 1434  
 Db 1410 LTAGPKLLEPCSPRIPILTAHLELQQA--EGIDAGQWQPAQSEBALERVENEPFGIGE 1468  
 QY 1435 EQTLTLKRLKDELARRELKSLGSESSGTTPVGSFHTEA--RWT---DYSLSR--- 1486  
 Db 1469 PE-----GLQDWEEGEESSEADLLETLPDSTPLGLYLSPASPKMDLAGEOQLSPQD 1522  
 QY 1487 -----PAKEALASDSQNGOEGSCPEBESDIA---LEDSADTDTANSPG 1526

Db 1523 AGKEDWGPVAPAAQGLSGPPEEEBEG-----HGSJDSSEEFEDLGTASILLPG 1571  
 RESULT 4  
 US-09-079-812E-2  
 ; Sequence 2, Application US/09079812E  
 ; Patent No. 6340575  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Bollag, Gideon  
 ; APPLICANT: Crompton, Anne  
 ; APPLICANT: No. 6340575th, Anne  
 ; APPLICANT: Sharma, Sanju  
 ; APPLICANT: Roscoe, William  
 ; TITLE OF INVENTION: Methods and Compositions for Treating Abnormal Cell  
 ; TITLE OF INVENTION: Growth Related to Unwanted Quanine Nucleotide Exchange  
 ; FILE REFERENCE: 1028-US  
 ; CURRENT APPLICATION NUMBER: US/09/079, 812E  
 ; CURRENT FILING DATE: 1998-05-15  
 ; PRIOR APPLICATION NUMBER: 60/049, 879  
 ; PRIOR FILING DATE: 1997-06-17  
 ; NUMBER OF SEQ ID NOS: 33  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO: 2  
 ; LENGTH: 710  
 ; TYPE: PRT  
 ; ORGANISM: Layer Rac GEP  
 US-09-079-812E-2  
 Query Match 3.3%; Score 258.5; DB 4; Length 710;  
 Best Local Similarity 21.3%; Pred. No. 4,3e-11;  
 Matches 146; Conservative 120; Mismatches 270; Indels 149; Gaps 29;  
 QY 506 ESSRSCTAEKTQASAPDKW-----LPPFKTRKQSSNKKREDALEDK----- 549  
 Db 2 ETRESEDELEKTRKKSASDQONTNENPAKYPELLPEKEETSQADQ---IQDEPHCHP 58  
 QY 550 -KRNPL-RYI--GKPRSSQSIRGNVRNIIQHNESHQYDV----- 589  
 Db 59 IKRNSIFNRSIRKSKAKARDNPERNASCLADSQDNKSVNEPLTINIPWSRMPORITAM 118  
 QY 590 --EPGTORLS-----TGSFPED---LLESDSSRSEI-----RGRGSGSLK 624  
 Db 119 QTPDGAQEMSSSTPGNGATPEWPAALDPTTLTEALRMHP IPADSQRNILEIGLL 178  
 QY 625 GREEMKSRKAENVPRPRSVDMDMAAFAARLHQSSASSSASTRSLNPTPTPKMG 684  
 Db 179 YQERYRDKSTQLELTRQODAEIEDNTNGS-----PASQDPEEEBEEBEEBEPASPE-- 232  
 QY 685 RRSIESPNLGECTDVILPHL---LEDDLQSLDLEPPEVQNMQHTVGKDVANLTQREI 741  
 Db 233 RKTLL--POLCLLSN---PHSRENIIMODL-----PETRS-----SGVLEIILQPEET 272  
 QY 742 DREGVINELFVTASHLRTLVLDLIFYQMKRENLMPRELARLPNLPBELIEHNSMC 801  
 Db 273 KLOEAMFEIYTSBASYKSLNLLVSHPMENERIRKTLIHSEAHILFSNLDVLAVEREL 332  
 QY 802 EAMKKTLEEGPILTRDISDPAARFADGPAREELQOVAOFCYSOVALELRTKQRESRP 861  
 Db 333 LEIHEHMEENIYVSDVC--IVRY---ADHF--SVITTVYSNTOYBERITYKOLLOKKA 387  
 QY 862 QLEMOAESHPQCRRLQDLNLDLIVSEMOQLTKYLLLENITKHEGTSHEKICLRARDQ 921  
 Db 388 RELIAQLELDPPKRCGRPFSSFFILPQRTIRLKLIVONILKRYE---ERSEERCTALDN 444  
 QY 922 RE---ILKFVNEAVKOTENRHRLEGYOKRLDATALERASNPAAEFKSLDLTRKKNIHG 978  
 Db 445 KELEMYVAKACNEGVKMSRTQOMISIQKME---FKIKVPIIISHSRWLKQGOELOQMSG 501  
 QY 979 PLTWIRISKDCTL--DLOVLLLEDVLVULOR-----QEBRLIKCHSKTAVGSSSDS 1026  
 Db 502 PKTSRLTKRKLKHEIYLFLENDLIVICRQIPQDKQVQDSAPRGLLRBE-----LEDO 556

```

Oy 1027 KQESPVLKINAVIRSV--ATKRAFFICIS-----ELGPOIIEVAIASSD 1074
      :::::|:::|:::|:::|:::|:::|
Db 557 GQTLA-----NVLFLRLLEAADDREATYIAKSSQSEMKRMWTSLAPNRKIFVSFSIR- 610
      :::::|:::|:::|:::|:::|:::|
Oy 1075 KNIMELLEEAIVONATKHGCAPIP 1099
      -L-|D-|C-P|O-V-C|H-P|Y-A|O-P 628
Db 611 -----L|D|C-P|O-V-C|H-P|Y-A|O-P 628
      -L-|D-|C-P|O-V-C|H-P|Y-A|O-P

```

```

1      RESULT      5
2      ; Sequence 4, Application US/07853913
3      ; Patent No. 5338839
4      ; GENERAL INFORMATION:
5      ; APPLICANT: McKay, Ronald D.G.
6      ; APPLICANT: Lendahl, Urban
7      ; TITLE OF INVENTION: Nestin Expression As An Indicator of
8      ; TITLE OF INVENTION: Neuroepithelial Tumors
9      ; NUMBER OF SEQUENCES: 4
10     ; CORRESPONDENCE ADDRESS:
11     ; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
12     ; STREET: Two Millitia Drive
13     ; CITY: Lexington
14     ; STATE: Massachusetts
15     ; COUNTRY: U.S.A.
16     ; ZIP: 02173
17     ; COMPUTER READABLE FORM:
18     ; MEDIUM TYPE: Floppy disk
19     ; COMPUTER: IBM PC compatible
20     ; OPERATING SYSTEM: PC-DOS/MS-DOS
21     ; SOFTWARE: Patentln Release #1.0, Version #1.25
22     ; CURRENT APPLICATION DATA:
23     ; APPLICATION NUMBER: US/07/853,913
24     ; FILING DATE: 19920319
25     ; CLASSIFICATION: 435
26     ; PRIOR APPLICATION DATA:
27     ; APPLICATION NUMBER: US 07/660,412
28     ; FILING DATE: 22-FEB-1991
29     ; PRIOR APPLICATION DATA:
30     ; APPLICATION NUMBER: US 07/603,803
31     ; FILING DATE: 25-OCT-1990
32     ; PRIOR APPLICATION DATA:
33     ; APPLICATION NUMBER: US 07/201,762
34     ; FILING DATE: 02-JUN-1988
35     ; PRIOR APPLICATION DATA:
36     ; APPLICATION NUMBER: US 07/180,548
37     ; ATTORNEY/AGENT INFORMATION:
38     ; NAME: Granahan, Patricia
39     ; REGISTRATION NUMBER: 32,227
40     ; REFERENCE/DOCKET NUMBER: MIT-4641AAAA
41     ; TELECOMMUNICATION INFORMATION:
42     ; TELEPHONE: 617-861-6240
43     ; TELEFAX: 617-861-9540
44     ; INFORMATION FOR SEQ ID NO: 4:
45     ; SEQUENCE CHARACTERISTICS:
46     ; LENGTH: 1618 amino acids
47     ; TYPE: AMINO ACID
48     ; TOPOLOGY: linear
49     ; MOLECULE TYPE: protein
50     ; US-07-853-913-4

```

Query Match	3.1%;	Score 247.5;	DB 1;	Length 1618;
Best Local Similarity	20.1%;	Pred. No. 1.1e-09;		
Matches 319;	Conservative 192;	Mismatches 539;	Indels 541;	Gaps 76

[illegible]

```

QY      1223  LRGHTVKIQAG----- : : : : : ||| | : -EP--EDDLTPPVSVITSHPW 1254
Db      1026  VEGSAVKGAGGLDDPPGCGSQOQVGTQLDAPQGLPEAIPLEYEDVAP----- GC 1075
QY      1255  DPGSP---GGAPTISDTRIAREEGSQPEGEDAVASSLAHLPPRTSSGVWSDBELDRN 1310
Db      1076  DASPEVMLGSEPMGBSASAGAERGLQGQVG----- GLDDPHLIRE 1117
QY      1311  PAAEASTPEPAASKYVR----- KVSLLREGGYGA--AKVAGSN---AIPDSQSSEEL 1359
Db      1118  EYMEPPLEESLEAKRVQGLEGRKLDLEEAGIGTGFSELPKSRDPWPMPREGEESBA 1177
QY      1360  SEVEGGAO-----TGACFVYVMPAGPLDSTET-PTGTPPSPOCHSLPAMP-TEP 1408
Db      1178  EAPRGAEEAFPAEILGHTGSDASPMPLSGEAEEDVPYLVSPSTYPPILEDAPGLQP 1237
QY      1409  QPYR-----GVRGCOCSSLVRRDVDYIFHTIEQL--TIKLRLKDMLAHRELKSLG 1459
Db      1238  QAEGSQEASMVCQ-----RAEGKVESQEDELGSCEIPEGLQDEEGSEREESEDEL 1290
QY      1460  GESSGCTTPVGSF--HTEARMTDY-SLSPAKE-----ALASDSQNQDEGSCPE 1507
Db      1291  GETLPDSTPLEFYLRSPTSRMPPLBSRGHPLKETKEGMDAVLASE---GLEPPSEKE 1347
QY      1508  EG-----SDIA--LEDSATDTRAVSPG 1526
Db      1348  EGEGEEECGRHSDLSFEFFDLGTENAPFLDG 1378

RESULT      6
US-08-466-390-4
; Sequence 4, Application US/08466390
; Patent No. 568562
GENERAL INFORMATION:
APPLICANT: TOKUATLY, GARY
APPLICANT: LIDBARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,390
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESQ, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7100
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-390-4

```





```

RESULT      8
US-08-467-781-4
; Sequence 4, Application US/08467781
; Patent No. 5780596
; GENERAL INFORMATION:
; APPLICANT: TOURKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/467,781
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ. ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2101 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-467-781-4

Query Match      3.0%; Score 239; DB 1; Length 2101;
Best Local Similarity 19.3%; Pred. No. 7, 2e-09;
Matches 239; Conservative 210; Mismatches 469; Indels 320; Gaps 56;

```

```

QY      621 GSLKREEMK-----RSRKAENVPRPSDYMDMAAEARLHOS 659
      : : : : :
      : : : : :
DB      440 ETERGOOEAKLMBRGHFEEREKQOQLSLTLDQSSISINLSOAKEEIEQASQAGARLTAQ 499
      : : : : :
QY      660 ASSASSLSTRSLNPPTPTPKMGRSISPLGCTDVTIPLHLEDLIGQSDL--EP 717
      : : : : :
DB      500 VASLTSELTTLN-----ATIQOQOELAG--LKOAKEQAOQLQTLQOQ 542
      : : : : :
QY      718 EPEYQWMOHTVGADVVAANTOREIDROEVINELFVTEASHLR-----TLRVLDL 766
      : : : : :
DB      543 EQASQGLRHQV-EQLSSSLKQFOQLKVAKEKEALRODHAQOLATAAREKRESLSKERD- 600
      : : : : :
QY      767 IFYQRMKKENLMPREELARLPMLPELIEIHN-----SWCEAMKUREEGPIITRDISDPN 821
      : : : : :
DB      601 ---ALKQLEALKEKRAAKL-ETLQOQOLQVANAEDBSAQTSYQAOERKAELSRYEE-- 654
      : : : : :
QY      822 LAFEDGPARELQOVAAQFSYQSVALTEL-IRKQKRESFQJFMQEAESHPOCRRLQLR 880
      : : : : :
DB      655 LQACVETARQEOHEAQOVAE-----LEIQLRSEQOKATEKEKVAOE-----K 697
      : : : : :
QY      881 DLIVSEMQRLTKYPLLENIIRKTEGTSSEHEKICRARDQREILKFVNEAVQOTEN--- 937
      : : : : :
DB      698 DQLOEQDQ-----ALKESLKYTKGSLSEERK--RAADALEQOQCSISELKAETRSIVE 748
      : : : : :
QY      938 ---RHRLEGYQKRLDATALERASNPILAERKSIDLITRKMIHGC----- 978
      : : : : :
DB      749 QHKRERKELEERAGRGKGEARLLQGEAHQAETEVLRRELAPAMAQHTAESECBQLYK 808
      : : : : :
QY      979 -PLTW---ISKDKTLDLOVLLLEDLYVL-----LQROERLL-LKCHSKTAVG 1022
      : : : : :
DB      809 EVANAMDGYDSOEERAOYGAMFOEQLMTLKECEKARQELQAKKAVAGIESHLSQTS 868
      : : : : :
QY      1023 SSDSKQTFSPVLKNAVLIRSV--ATDKRAFITTCSELGPPQIYELVALTSSDKNIME 1080
      : : : : :
DB      869 RQONK-----LAEIHLANLALQOQVOEKEVRAOKLADLSTLD--EKMATSKEE---VA 917
      : : : : :
QY      1081 LEEAVQONTKHKGAPIRP--HSPPGSGEPAY-----QGS---STSSRVEINDSEVYHT 1130
      : : : : :
DB      918 RLETLVRKAGEOQETASRELVKPEPARAGDRQPEWLEBQGRQFCSYQALQAMAREAEQM 977
      : : : : :
QY      1131 EKEPKKL-----PEGPGPEQROVQDKLI-----AGEPEVQO----- 1161
      : : : : :
DB      978 GNFLERLRALMESQOQOQOQERQOGERVARLQNGERGAQADALALKARAELMRLQNA 1037
      : : : : :
QY      1162 --EDEBELKTLPPRPSLDGENRCITRDPVLLATLGPILMBGLADALEVDENLHLIL 1219
      : : : : :
DB      1038 LNEQARVEFATLQELAHALTEKEG--KQOELAKL-----RGLAAQIKKELELQTV- 1087
      : : : : :
QY      1220 WSLLPQHTYVQAAGEPEDLTPTPSVSTSHPMQPGSGQAFTISDRRLARPRGSPQ 1279
      : : : : :
DB      1088 -----KOLKEOLAKKEKEHAS-----GSAQ-----SEAGRTPTGPXU 1122
      : : : : :
QY      1280 EGEDVAVVSLAHLPPRTSSGVDSPELDRNPAAEAAS 1317
      : : : : :
DB      1123 EALRAEVSKLEQOCQCKQOE-----QADSLERSLEAEAS 1156
      : : : : :

RESULT      9
US-08-195-487-4
; Sequence 4, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOURKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA

```

```

COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/195,487
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/901,701
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: PITCHER ESO, EDMUND R
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: MTP-013
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/248-7000
TELEFAX: 617/248-7100
INFORMATION FOR SRO ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2101 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-195-487-4

```

```

Query Match 3.0%; Score 239; DB 1; Length 2101;
Best Local Similarity 19.3%; Pred. No. 7.2e-09;
Matches 239; Conservative 210; Mismatches 469; Indels 320; Gaps 56;

```

```

QY 246 RFPSTSELMARNVSLPGDLSQTSPIYLARAQ-----HHRQSGDAALLP 294
DB 73 KHPSSPECLVSAQVLEGGSELELAKMTLLYHSTMSKSPRDWEQFEYKIQALAVILK 132
QY 295 --LNHOG-----IDSPKPLIIG--PEEDYDGYFNNESSDIIFODLEKLKSHP 338
DB 133 FVLDEHEDSLNEDLENLQKAPVSTGSSFPPELSPPSH-QAKREIRFLELQKVASSS 191
QY 339 AYLVLFLRYILSQADPGFLFLYLCSEVYQQTNPK---DSRSLGDIWNIFLEKNAPLRY 394
DB 192 SG-----NNFLSGSPASPMGILQTPQFQMRRLKQLADERS--NRDELELELAERKILLT 245
QY 395 KIP---EMLQAEIDLRLANNDDPRNVLCQAQAVMLEIQ--EQINDYKSKRTILGSLYIG 449
DB 246 EKDAQIAIMQORID-RLA-----LNEKQASPLEPEKELELRDNESLJMRL----- 292
QY 450 ENDLGLDGLDPLREROMAEKQALALGDLISXYEDRSAPMDFAVTEFMSHAGIRLESRS 509
DB 293 -HEFLKQOQODLKTESQMDRKINQL-----SEENDLSFKLRFPASHIQ-OLQDPLN 342
QY 510 SCTAEKTOSAPDKDKWLPFPFKTKKSSNSKKEKDALEDKK---RNPILRYIGKPKSSS 565
DB 343 ELTRESKAT---QEWL-----EKQALEKELSAIADKKLEKKEKNIILQ----- 384
QY 566 QSIRPGVNRNIIHFENSHQYDVEPG-----TQRLTSGSPEDULESDSSRSSEIRLGRS 620
DB 385 -----GKLSQLEHLSQLQDNPNPOKEGVLGDVQLQLETKQEAATLANNTQLDARVEML 439
QY 621 GSLKGREEMK-----RSRKAENVPRPSDVDDDAEAARLHOS 659
DB 440 ETERGQOQAKLLAERGHFEERKQSLITLQSSISLQAKELBASAHGARLRLAQ 499
QY 660 ASSASSLSTRSLENPPPTPKMGRSIESPNIGFCTDVILLPHILEDLQQLSGLD--EP 717
DB 500 VASLSTSELTTLN-----ATIQOQDDELALG--LKQAKKEKQALQOTLQOQ 542
QY 718 EPEVQNMHTGKDVVANLTQREIDROVINELPFTESHLR-----TLRVLDL 766
DB 543 EQASQGLKHQV-EQLSSSLKQKEQOLKEVAKQKATRODHAQQLATAEAEREASLURERD- 600

```

```

QY 767 IFYQMRKENLMPREELARLPNLPETLEIHN-----SWCEAMKLEEGPIIRDIDSPM 821
DB 601 ---AALKOLEALEKEKAAL-ELIQOOLQVANKENRDSAGSYTOAQKEKALSKEYE-- 654
QY 822 LAREDPAREELQOVAQFCYSQVALEL-TRTKQRESRQFLPMQAEHPQCRQLQLR 880
DB 655 LQACVETARQOHBAQOAVAE-----LELQRSQOKATEKEVERVAQ-----K 697
QY 881 DLIVSEMOURLTKYLLLENTIKKTEGTSHEKICLRARQOCRELLKTVENAVKOTEN--- 937
DB 698 DQLOEQLO-----ALKESLKVTKGSLSEER--RADALEEQORCISLAKTATSLVE 748
QY 938 ---HHRLEGYQKRLDQATLEASNPAAEFSLDLTKRMHGE----- 978
DB 749 QHKREKLEERAGRGKGLERLLQLGEMAOAETEVLRRELAEMAAOHTAESCEBOLV 808
QY 979 -PLTWR-----ISKDTIDLQVLLLEDVL-----LQROERLL-LKCHSTAVG 1022
DB 809 EVAAMRDQVEDSQOEQAQYGAMFQEQMLTKKECEKARQELQAKKQVAGIESHSELOIS 868
QY 1023 SSDSKQTFSPVLTAKNAVLRSV--ATDKRAFFITCTSELGPQIYELVALTSSKNTIME 1080
DB 869 RQQRK-----LAEHLANLARALQOVQKEVRAQKADLSTLQ--EKMAATSKK-----VA 917
QY 1081 LLEBAVQNAWKHPGAAPRPI--HSPSGQSEPAY-----STSSRYEINDESYHT 1130
DB 918 RLETLVRKAGQOQETASGLYKPERAGDRQPEWLEEQGQGFQSTQALQAMEREAQM 977
QY 1131 EKEPRKL-----PEGPQEQRVQDKOLI-----AGEPVQV----- 1161
DB 978 GNELEIRLALMESQGOQOEERGQOEREVARLTQERGRADLALFKARAELMRQNA 1037
QY 1162 --EDEEELRTLPRAPISLDGENRGIRTRDPVLLATGRLMEGLDADALEVEULRLIL 1219
DB 1038 LNEORVEPATQELBALHALLKEG--KDOELAKL-----RGLEAQIYELEELRTV- 1087
QY 1220 WSLPGLHTVKTQOAGEPEDDLTPTPSVYSITSHPDGSPQOAPTISDSTRLARPEGSSOP 1279
DB 1088 -----KQLKQALKKEKEHNS-----GSGAQ-----SEAGRTPTGPKL 1122
QY 1280 EGEDVAVNSLAHLPRTRRSSGVWDSPELDNRNPAEAS 1317
DB 1123 EALRAEVSKLEQOCOKQOE-----QADSLERSLEAERAS 1156

RESULT 10
US-08-483-924-4
Sequence 4, Application US/08483924
Patent No. 5882876
GENERAL INFORMATION:
APPLICANT: TONKATLY, GARY
APPLICANT: LIOGARD, GRAHAM P
TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
NUMBER OF INVENTION: 6
TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
CORRESPONDENCE ADDRESS:
ADDRESSEE: TESTA, HURWITZ & THIBEAULT
STREET: 125 HIGH STREET
CITY: BOSTON
STATE: MA
COUNTRY: USA
ZIP: 02110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,924
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

```

```

: NAME: PITCHER, ESQ., EDWARD R
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: MTJ-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 248-7000
: TELEFAX: (617) 248-7100
: INFORMATION FOR SEQ. ID NO.: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2101 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: OS-08-483-924-4

```

Query Match	3.08;	Score 239;	DB 2;	length 2101;
Best Local Similarly	19.38;	Pred. No. 7.2e-09;		
Matches 239;	Conservative 210;	Mismatches 469;	Indels 320;	Gaps 56

QY	246	RFPSTSELMNRNVSJLSDPGJDSPTSPILARVAQ-----HHRGSGDALLP	294
Db	73	KHPSEPVELVSAQKYLEGESELELAMTMTLLHYSTMSSKSPPDMOEFEYKIDAEVLVILK	132
QY	295	--LNHG-----IDOSPRLIIG---PEEDYPGYFNNSDILFOOLEKLKSHP	338
Db	133	FVLBHEDELINNEDELENFLQKAPRVSTOCSSTPFEEISPSH-QAKREITFLELOKVASS	191
QY	339	AYLVLFELRYILSQADPGPLLYGCSEVYOQTNPK----DSRSLGDIMWIFLEKANPLRV	394
Db	192	SG-----NNFLSGSPASPMGDILTQTPQOMRKQLADERS-NRDELELELAENKLLT	245
QY	395	KIP---EMLQAEIDLRLNNEBPNAVLCQAQAVMLEIQ-BOINDYRKRTLGSLG	449
Db	246	EKDQAIAAMQORID-RILA-----LNEKQAASPLEPELELRKNKNESTMR-L----	292
QY	450	ENDLGLDGDPLEBRQMAEKOLAALGDILSKYEEDRSAMDVAVTWMFHAGIRLEERS	509
Db	293	-HEITLKQCODIKTEKSQMDRKINOL-----SENGDLSTKLKEFASHLO-QUDALIN	342
QY	510	SCTAETKOASAPDKDWLPFFPKTKROSSNKKEKDALEDK--RNPIARYIGRPSSS	565
Db	343	ELTEHSKAT--QEWL-----EKQAOLEKRLSALOQKCKLEKENILO-----	384
QY	566	QSIRGNVRNIIOHFENSHOIDVPERG-----TORLSTGSFPEDLLESOSSSEIRLGNS	620
Db	385	-----GKLSOLEEHLISOLDNPNPOEKGEGVLDVLOLETILKQEAATLANNTOLOAREVM	439
QY	621	GSLGREEMK-----RSRAENVPRRSVDMDAAALEARLIHS	659
Db	440	ETERQOQAKLLAERGHFEEEKOOLSILTDLOSSISNLQAKEELEQASOHGAKLTQAQ	499
QY	660	AASSASSSTRSLENPPPTPKMGARSIESPNLFCFDVILPHULEDDLOLSDL-EP	717
Db	500	VASLSTELLTLN-----AITQQDDGLAG--LKQAQKEQAQALQTIQOO	542
QY	718	EPEYONMOTHYGKDVANLTOREIDROEVINELFYTEASHLR-----TLRVLDL	766
Db	543	EOASCGELTHOV-EQUSSSLKQEKQOOLKVEAQEATRODHQAOLTTAAEEERASLERPD-	600
QY	767	IFYQRMRENMPRELARLFPNLPBELLEINH-----SMCEMKYLREBGRIIROISDM	821
Db	601	--AALKQALELEKKAKL-ELLQOQLOVANEARDSOTSOTQAORAEALSERVEE--	654
QY	822	LARFDGPAREELOQVAAOFCSYOSVALEL-IRTKORKESRFOLFQOEAESHPOCRRLQR	880
Db	655	LQACEVTAHQOEHQAOVAE-----LELOLRSEQOKATEKERVAOE-----K	697
QY	881	DLIVSEMORTLKYPLLBNIIKHTEGTSEHEKLCRARDOCREILKFVNAVKQTEN--	937
Db	698	DQLOEQOQ-----ALKESLKVTKGSLJEERK--RAADALEEQGRCSSELKAEFRSLVE	748
QY	938	---RHREGVOKRLDATALEBASNDLAAEFKSLDITTKMHIEG-----	978

```

Db 749 QHKRRKRLLEEBRGRGKGLERLLQJGEAQAQETVLRNLEMAAQAHTASBEOQLVK 808
Qy 979 -PLTW- ----ISKDTIDLOVLLLEDLVL-----LOROEERLL-LKCHSKTAVG 10222
      ||
      ||
Db 809 EYAAARCDVEDSDOEAEAOYCAEMEOJMLTKECECKAROELOEAKKEVAGISHSHELQIS 868
Qy 1023 SSDSQMOTFSPLLKLTANVLIRSV--AUDPRAFLICHSGLPPOIVELVATVSSDKINIMME 10806
      .:|
      .:|
Db 869 ROONK-----IAELHMANIARLQAOVDEKEVYRAOKLADDLSTLO--EKMANTKE-----VA 917
      .:|
      .:|
Qy 1081 LLEAVONATKHPGAAPIPI--HSPPGSOEPAY-----OG-----STSSRVELINSEVYHT 11300
      .:|
      .:|
Db 918 RLETLVRKAGBOQETASRELIVKEPARAGDRQPEMLEEOGRGCSTQALQAMEAREAMQ 977
      .:|
      .:|
Qy 1131 EKEPKKL-----PEGGPGPQROQDKOLI-----AQGEVQOE----- 11616
      .:|
      .:|
Db 978 GNEIERLIALMESOGGOQOERGOQREYARLTOERGRQADDALEKKAARAELEKRLONA 10377
Qy 1162 --EDEEELRTLPRAPPSLIDENRNGIRTRDPVLLATLGPRLMEGLADDALEDEVENRHLIL 12199
      .:|
      .:|
Db 1038 LNEQVREVAETIQEALNAHALTEKES--KQDELAKL-----RSLBAEQIKLEIELRQVIV- 10875
      .:|
      .:|
Qy 1220 WSLPLGHTVKTQAAGEPDDLTTPPSVVSITSHPMWPGSEFGQAPTISDSTRLARPEGGSP 12799
      .:|
      .:|
Db 1088 -----KOLKEQLKAKKEKEHNAS-----GSGAQ-----SEAGRTPTPTGPKL 11222
      .:|
      .:|
Qy 1280 EGEDVAASLHLBPRTRSSGVMDSPELDRNPAAEAS 1317
      .:|
      .:|
Db 1123 EALRAEYSKLEQOCQOKQOE-----QADSLENSLEAEKAS 1156

```

## RESULT 11

US-09-452-294-1  
; Sequence 1, Application US/09452294

GENERAL INFORMATION:

APPLICANT: Bissell, Mina

TITLE OF INVENTION:	DIFFERENTIATION DISORDERS
TITLE OF INVENTION:	THERAPY AND DETECTION OF PROLIFERATIVE AND
TITLE OF INVENTION:	

Patent No. 6287790

; CURRENT FILI

; PRIOR FILING DATE: 1998-11-30

```

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1

```

```

; LENGTH: 2101
; TYPE: PRT
;

```

ORGANISM: HOMO SAPIENS  
US-09-452-294-1

Best Local Sim

00 346 PENDING

73 KHPSSPECTVSAOKVT

QY 295 --LNHOC-----

Db 133 FVLDD

	Query Match	3.0%	Score 239;	DB 4;	Length 2101;	
	Best Local Similarity	19.3%;	Fred. No. 7.2e-09;			
	Matches 239;	Conservative 210;	Mismatches 469;	Indels 320;	Gaps 56;	
OY	246	REPSISESLMNNNSVLSPOGLDPSPTSPVIILAVNAQ-----HHRROGSDAALLP	294			
	:	:  :  :	:  :  :	:	:	:
Dd	73	KHPSSPECTVSAOKVLBESSELELAKMTMLLLHYSTMSKSPRDMEQFEYKIQALAVLT	132			
OY	235	--LNHG-----IDOSPRLITG---PEEDYDPGFNNESEDIIIFODLEKLAKHP	338			
	:	:  :	:  :  :	:	:	:
Dd	133	PVLDEHDGDLINTEDELNFLOKAPVASTCSSTFPPELSPSH-QAKRETRFLELOKVASSS	191			
OY	339	AIVLVFLATYLIQAADPGCLLFYLCEVYGQQTNPK----DSRSKGDMINFIIEKANPLRV	394			
	:	:  :  :	:  :  :	:	:	:
Dd	132	SG-----NNFLSGSPASPMGDILTQPPQFMRRLLKQLADLERS-NRDELELELAENRKLLT	245			
OY	395	KIP---EMIAOAEIDLRLFNNEDEPRNVLCLEAOEAVALEIQ--EQIINDYSKRRTLGISLYG	449			

```

Db      246 EKDAQIAMQORID-RLA-----LINEQAASPLEPEKLELDKDNESLTMRL----- 292
QY      450 ENDLLGLDDPLREROMAKOLAALGDLISKYEDRSAPMDFAVNTFMSHAGIRLRESRS 509
Db      293 -HETLKOCODLTKTESQMKRINQL-----SEENGDLSEFKLEFASHLO-QLQDALN 342
QY      510 SCTAEKQASAPDKKWLPPFPKTKKQSSNKKEDALEDK-----RNPILRYIGKPKSSS 565
Db      343 ELTEHESKAT---QEWL-----EKQAOLKEKLSAALQDKKCLEKKNELIQ----- 384
QY      566 OSIKGNVNRNIIOHFNESHQYVPEPG-----TQRLSTGSFPEDLLESDSSRSRIRLGRS 620
Db      385 -----GKLSQLEHLSQLQDNPPQEKGEVLGDVLIQLETLKQEAATLAANNTOQLQARVEML 439
QY      621 GSKLGREEMK-----RSKKAENVPRPSVDVMDAAAEARLHOS 659
Db      440 ETERQOQEKALLERGHFEERKQOLSLITDLOSSISNLSQAKKELEQAQSHGRRLTAQ 499
QY      660 ASSASASLSTRSLBNPTPPTPKMGRSIESPNLGFCTDVLPHLELDLQGLSDL-EP 717
Db      500 VASLISELTLTN-----ATIQODQDELAG---LKQAQKQQAOLAQTLQOQ 542
QY      718 BEPVONMOHTVGDVYANITQREIDROEVINELFVTEASHLR-----TLRLVLDL 766
Db      543 EQASQGLRHQV-EOLSSSLKQKQOOLKEVAKEQKQATRODHAQOLATAAEEREASLRERD- 600
QY      767 IFYQNRKKNLMPREELAFPNLPELLEIHN-----SMCEMKMLKREGPIIRISDPM 821
Db      601 ---AAIKQLEALEKEKAAL-ELIQQQLQVANEADSAQTSVTAQORKEAELSRVEE- 654
QY      822 IARFDGPARRELQOYAAQFCSYQSAVLEL-ITKQKRESFQOLFQQAESHPOCHRLQJR 880
Db      655 IQACVETARQOHEQAQVAE-----LEIQLRSEQQKATEKERVAAQ-----K 697
QY      881 DLIVSEMQRLLTKYPLLENIITHTGEGTSEHKLCRAPQCEILIKFVNEAVKQEN- 937
Db      698 DQLOLOLQ-----ALKESLKVTKGSLSEERK--RAADALEEQORCISLKAETRSIVE 748
QY      938 ---RHRLGYQKRLDATALERASNPLAEBFKSLDLTKRMHIEG----- 978
Db      749 QKRRERKELEERBARAKGILEALDLOGEAHQAFVLRRELAEMAAOHTAESCEQOLVK 808
QY      979 -PLTWR---ISKDKTLQVLLLELDVYL-----LQROBERLL-LKCHSKTAVG 1022
Db      809 EVAAMRDQGEVDSQOEAAQYGAWFQSOQLMTLKECEKAKQELQEAKEVAGIESHSLDIS 868
QY      1023 SSDSKQTSPLVLAQVILRSY--ATDKRAFITCTSELGPPQIYELVALTSSDKNIME 1080
Db      869 RQONK-----LAELHANLARALOQVQKEVRAQKLDADLSTLQ--EKMAATSK-VA 917
QY      1081 LLEEYONATKHPGAAPIPi--HPSPGSOEPAY-----OG-----STTSRVINSEYVHT 1130
Db      918 RLETLVYRAAGEQOETASSELVYKEPARAGDROPEWLEBOQGFQFQSTQALQAMEREABQM 977
QY      1131 EKEPKKL-----PEGPPEQORVODKOLI-----AQGEPEVQOE----- 1161
Db      978 GNELERLAALMESQGOQOERQOERREYARLTQERGRAQADLALEKARAARLEMLRQNA 1037
QY      1162 --EDBEELRTLPRAPPSLDGENRGIRTRDPVYLAITGLPLMEGLADALEVDENIRHLITL 1219
Db      1038 LNEORVEEATLOELAHALHTEKEG---KDOELAKL-----RGLEAAQIKLELEELRQTV- 1087
QY      1220 WSLLEPHVTQOAGEPBDLTPPVSIVTISHMPDQSPQOAPITISSTRLAREGSGOP 1279
Db      1088 -----KOLKEQLAKKEKEHNS-----GSGAQ-----SEAGRTPEPTGKXL 1122
QY      1280 EGEDVAVSSLAHLPPRTSSSGVWDSPELDRNPAEAAS 1317
Db      1123 EALRAEVSKLEQOCCQOQE-----QADSLERSLEABERAS 1156

```

```

PCT-US93-06160-4
: Sequence 4, Application PC/TUS9306160
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
: TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
: NUMBER OF SEQUENCES: 6
: CORRESPONDENCE ADDRESS:
: ADDRESS: TESTA HURWITZ & THIBEAULT
: STREET: 53 STATE STREET
: CITY: BOSTON
: STATE: MA
: COUNTRY: USA
: ZIP: 02109
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/06160
: FILING DATE: 19930621
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: PITCHER ESO, EDMUND R
: REGISTRATION NUMBER: 27,829
: REFERENCE/DOCKET NUMBER: MFP-013
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617/248-7100
: TELEFAX: 617/248-7100
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2101 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: PCT-US93-06160-4

Query Match      3.0%; Score 239; DB 5; Length 2101;
Best Local Similarity 19.38; Pred. No. 7.2e-09;
Matches 239; Conservative 210; Mismatches 469; Indels 320; Gaps 56;

QY      246 RFPSTESIMRNRSVLSPPGIDSPQTSPLIARVAQ-----HHRQSGDALLP 294
Db      73 KHPSPCELVSAQVYLBESSELELAKMTLLYHSTMSSKSPROMQEPYKIQALAVILK 132
QY      295 --LNHQG-----IDQSPKPLIIG--PREYDPGYFNESDILIFQDLEKLSHP 338
Db      133 FVLHDEDLINEDLENFLQKAPVPSTCSSTFPELSPSH-QAKREIRFLEIQVASS 191
QY      339 AYLAVFLAYILISQADPGPLFLYLCSEVYQQTNP-----DSKSLCKDLMNIFLEKNAPLRY 394
Db      192 SG-----NNFISGSPASMGDILOTPOFMRLKQOLADERS-NRDELLELEENRLLT 245
QY      395 KIP---EWLQAEIDLRLRNNEPDRNVLCQEAQVYMLEIQ--EQINDYRSKRTGLIGSLYG 449
Db      246 EKDAQIAMQORID-RLA-----LINEQAASPLEPEKLELDKDNESLTMRL----- 292
QY      450 ENDLLGLDDPLREROMAKOLAALGDLISKYEDRSAPMDFAVNTFMSHAGIRLRESRS 509
Db      293 -HETLKOCODLTKTESQMKRINQL-----SEENGDLSEFKLEFASHLO-QLQDALN 342
QY      510 SCTAEKQASAPDKKWLPPFPKTKKQSSNKKEDALEDK-----RNPILRYIGKPKSSS 565
Db      343 ELTEHESKAT---QEWL-----EKQAOLKEKLSAALQDKKCLEKKNELIQ----- 384
QY      566 OSIKGNVNRNIIOHFNESHQYVPEPG-----TQRLSTGSFPEDLLESDSSRSRIRLGRS 620
Db      385 -----GKLSQLEHLSQLQDNPPQEKGEVLGDVLIQLETLKQEAATLAANNTOQLQARVEML 439
QY      621 GSKLGREEMK-----RSKKAENVPRPSVDVMDAAAEARLHOS 659

```

```

Db 440 ETERGOEAKILABRGHFEERKQOLSSLIJDQSSISNLSQAKELQOASAHGARTLAQ 499
QY 660 ASSASSLSRSLNPPPTPKMGRSIESPNIGFCTDVLPHLLDDGQLSDL--EP 717
Db 500 VASITSELTILN-----ATIQOQDOELAG--LKQAAKEKAQALQOTQOQ 542
QY 718 EPEVONNOHVYKGDVYANLQREIDROEVLNELLVTEASHR-----TLRVLDL 766
Db 543 EQASQGLRHQV-EQLSSGLKQEQOLKEVAKQEQATRODHQQLATPAERERASLRERD- 600
QY 767 IFYORMKRENIMPREELARLPNLPPELIEIHN-----SMCEAMKRLREEGFIINDSDPM 821
Db 601 ---AALQLELEKEKAKKL-ELIQOQLOVANEARDQAQSVTQAQKEKALSKVVE-- 654
QY 822 LARFDGAREBLOQVAAQFCYSQVSALEL-IRTKQKRESRQFLQMOAESHPOCRILQLR 880
Db 655 LQACVETARQOHEQAQVAE-----LEQLRSEQOKATEKEVEVAE-----K 697
QY 881 DLVSEMQRLTKYPLLENIKKHTEGTSHEKLCRARQOCRELKRVNEAVKQTEN--- 937
Db 698 DQLOEQLO-----AKESLTKVTKGSLKEER--RADALLEQOQRCISELKATRSLVE 748
QY 938 ---RHLEGYOKRLDATALERASNPPLAEFKSLDTTRKMIHEG-----978
Db 749 QHKEREKLEBERGRKGLERLLQGEAQAQETEVLRRELAEMAAQHTAESCEQOLVK 808
QY 979 -PLTW---ISKDKTLDLOVLIEDLVVL-----LQROEBRL-LKCHSKTAVG 1022
Db 809 EVAMRQDGYEDSOEEAQAQYAMFEQOLMTLKECEKARQELQEAKEKRVAGIESHSELOIS 868
QY 1023 SSDKQKFFSVLKLNAVLTISV--ATDKRAFFICTSELGPPQVLEYALTSSPKNTWME 1080
Db 869 RQOKK-----LAEIHAHARALQOQVEKVRQAOKLADLSTLO--EKMAATSK--VA 917
QY 1081 LLEAVONATKHGAPAPIPI--HPSPPGSOEPAY---QG---STSSRVEINDSEVYHT 1130
Db 918 RLETLVAKAGEQOETASRELVKEPARAGDQPEWLEQOQGFQSTQALQAMERAEQM 977
QY 1131 EKEPKL-----PEGGPQORVODKOLI-----AGEPQVE-----1161
Db 978 GNELEERLARALMESQGOQOBERGQEREVARLQOERGRAGADLALAEARAEEMRLQNA 1037
QY 1162 --EDEELRTLPRAPPSLDGENRGIRTRDPVLLATGPLMEGLADALADVENTLRHLIT 1219
Db 1038 LNEQRFVETALQERLALALEKES--KQDELAKL-----RGLLEAQITELERQTV- 1087
QY 1220 WSLPGHTVKTQAAGEPEDDLPPTPVYSITSHWPDGSPGQAPTTISDSTRLARPEGSSP 1279
Db 1088 -----KOLKEQLAKKEKEHAS-----GSGAQ-----SEAGRTPEPTPKL 1122
QY 1280 EGEDVAVSSLAHLPPTRRSSGWDSPELDNRNPAEAAS 1317
Db 1123 EALRAEVSKLQOQOQOE---QADSLERSLEAERAS 1156

RESULT 13
US-09-155-770-7
; Sequence 7, Application US/09155770A
; Patent No. 6300484
; GENERAL INFORMATION:
; APPLICANT: DuPont, David
; TITLE OF INVENTION: DNA ENCODING DP-75 AND A PROCESS FOR ITS USE
; FILE REFERENCE: 200130.418
; CURRENT APPLICATION NUMBER: US/09/155,770A
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 626
; TYPE: PRN
; ORGANISM: Homo sapien
US-09-155-770-7

```

```

Query Match 3.0%; Score 237; DB 4: Length 626;
Best Local Similarity 20.2%; Pred. No. 1,5e-09;
Matches 106; Conservative 106; Mismatches 219; Indels 94; Gaps 18;

QY 740 EIDR-OEVINELFTEASHRPTLRVLDLIFYORMKRNLMPREELARLPNLPPELIEIHN 798
Db 21 DADRKRKVIQELVDTEISYKDSLCELFELVLEPLQNTFTLQDBMESLFESLPEMLPEFK 80
QY 799 SWCEAMKRLREEGFIINDSDPMLARPDGARELOQVAAQFCSY-----QSV 846
Db 81 VPLETL-----EDG--IASSDPFTLETPSOFRKLFLSGLSGFLYYADHFKLYGSGFCANH 134
QY 847 ALELIRTKQKRESRQFLQMOAESHPOCRILQRLDLYVSEMQRLTKYPLLENIKKHTEG 906
Db 135 KQVKVLERAKTDQAFKAFI--DARNPTQHSSTLESYLIKRVQVRLKPKLLKEIVSLTD 193
QY 907 GTSHEKLCRARQOCRELKRVNEAVQQTNRHRLBEGYQRLDATALERASNPPLAEFKS 966
Db 194 ESEBHNLTLALAMEKAVASHINEMQKTYED-----YGVFPQOLVAEQS--GTEKEY 243
QY 967 LDLTRKMIHEGPLTW-----RISKDKTLDLOVLIEDLVVLQROEBRL--LLKCH 1016
Db 244 TELSMGELMHSTVSMNLNPLSLGKARKD--LELTVEVFRAVILVYKCKKCLKKLPSN 301
QY 1017 SKTAVGSSD-----SKQTFSVLKLNAVLTISVATDKRAE--ITCTSEL--GPPQIYELV 1068
Db 302 SRPAHNSSTDLPKFRMLIPISALQVRLGNPACTENNISWELIHTKSELGRCETTFOLC 361
QY 1069 ALTSSDKINIMELLEAV--ONATKH-----PGAAPPIHPSPGSOE-- 1109
Db 362 CSDESEKTNIVKAYIRSLIRENFRHICELPLEKTCQDRVLKKNVPPYAKLASRSKL 421
QY 1110 -----PAYQSTSSRVEINDSEVYHTEKEPKLPPEGPEQORVOD 1149
Db 422 VLKSSSEWETGTGKTLTLDSDGSLSGTQSGCPTAGRODSKSTSPGKYPHGLAD 481
QY 1150 --KQLAAGEPVOEDEELRTLPRAPPSLDGENRGIRTR---DP 1189
Db 482 FADNLIKESDILSDEDDHRQTVKQGPSPKMDIEIQFQRLISBDP 526

RESULT 14
US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.,
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: 312-474-6300  
 TELEFAX: 312-474-0448  
 TELEX: 25-3856  
 INFORMATION FOR SEQ ID NO: 5:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1780 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-769-309A-5

Query Match 2.6%; Score 203.5; DB 1; Length 1780;  
 Best Local Similarity 17.0%; Pred. No. 2.7e-06;  
 Matches 312; Conservative 259; Mismatches 704; Indels 557; Gaps 73;

29 SLSLSLGGSTPRTSPSHHROPSTSETTAGLVORCVIIQKDQHGFGTVSGDRIVLVOS 88  
 150 SSESMLLELT-----OPTESQANDIGFKKVFKEV-----GFKFTVKDKTEKEDPT 194  
 89 VR-----PGGAMKAGVKEGD-----RIIKVGTWVNTSSHLEYVKL 125  
 195 VOLLTVKNDGEGAGAGAGHODPSLAGEBAKSEPKOSTKPEETLKRESHAET--- 251  
 126 IKSAGYAALTLLGSSPPS-----VGVSGLQNPVAGVLRVNPILPPPPPPPL 174  
 252 -----SPPAESGQAVECKEKEGKEKESKSA-----ESPT 284  
 175 PPROHITGPKPLDPPEVQKHAHQ--ILW--NMLROEEELQDILLPCGGETSQTGGR 228  
 285 SPVYSETG-----STFKKFTQGWAGMKKTSFRKPKEDVE-----ASERK 326  
 229 LSVDSQADSGIDSGTEREPSSISESLMNRNSVLSDPGLDPSQTSVILARVAQHHRGGS 288  
 327 KQOEPEKVDTEEDGKAE-----VASEKLASGEGAHQGEAESHEERLAE 372  
 289 -DALLPLNHQ-----GIDOSPRLIIGPEEDYDPGFNNESDIIFO----- 329  
 373 YEKVELPEEEQVSGSQGSEKPAFL--ATEVEFDEKIEVHDEEVAAVHVSTVEERTEE 429  
 330 ---DLEKXL-SHPAYLVFLRITLSQADPGPLFLYLCSEVVOQTPKRSRLGKINWIF 385  
 430 OKTEVEETAGSVPAEELVGMADPOEAEPKELVKLETQVSGEDPTGADSLSPD- 484  
 386 LEKNAPLRYKLIPEMLQAEIDL-----RLRNNEPRLNLCQEAQVMLEIQEINDYRSKR 440  
 485 -EK--VLSKPEGVSEVEMLSQERMKVQCSPLKLIFTSLGKLKLGKQ--KGRK 536  
 441 TLGLGSLYENDLGLDGPLERQMAE-----KOLAALGDILSKYEEDRSAPMDF 491  
 537 G-GGDEESGEHTQVPADSPDSQEQKSSASSPEPEEITCLEKGLAEVQOD----- 588  
 432 AVNTKMSHAGIRLRSSSCRAETQOSAPDKDKWLPFFPKTKKSSNKKKEKDALEDKR 551  
 589 -----GEAEGATSDGKREKREGVTPMAF--KKMTTPKRVARRPSESDE 631  
 552 NPILEYIKPKSSQSIKPGVNRNIIIOFHENSQYDVEPTGORTOSTGSPFEDLESDS 611  
 632 DE---LKVKSATLSTESTASTASEMOEEMKGVPEPKPEEKRAKVDIVSMALICVSS 687  
 612 RSEIRLGRSGSLKGREEMKRS--RKAENVPRPSRDVMDAAEAARLHQSSASSAS- 665  
 688 KKRARRRSSDEBEGPRKAMGDKADADAKDK-ETGTGDLIAGSQEHDPGGSSSPQA 746  
 666 -----SLST-----KSLENPTPTPKMGRSISPMLGCTDVIPLHLEDLG- 710  
 747 GSPTEGEGVSTWSEFKRLVTPRKSSKSLKEKS-----EDSIAGSGV 788  
 711 --QLSDLEPEPEVONMHTVGDVYVANLTQREIDROE--VINELFVTEASH--LRTLVL 764  
 789 EHSITPDTPEGAE-BSWV-SIKKFTIGRRKKRDKQKQBDAPVEDGAPTANEDSDVPAVV 846  
 765 DLIFQMRKKNLMPREELARLPMLPELIEIHSNWCAMKRLREE----- 810

847 PLSEDAVEREKMEAQ--AQKGAQPE-----QKAAETVSELSSESYHMAAAVADGTR 900  
 811 -GPITRIDSDPMLARFQDPAREELQVAAQ--FCSYQVAALELIRTK-----QRK 857  
 901 AATIIERSPSMIS--ASVTEPLEQVBAEALLTEEVLEREVIAAEPPVTEPLEPENR 957  
 858 ESRFQLFMOEASHP-----QCRRLQDLVYSEMQRLTKPULLEN 899  
 958 EARGDTVSEALDTPAEVTAETGAPGLGSEGETSAAETTEMVASVQSULTDSPDTEE 1017  
 900 I--IKHTEGTSSEHEKLCRARDQCEILKFVNEAVK-----QTENRHLE 942  
 1018 ATPQVEVGVPDIEDQGR---RQGEVLQNAEKVKESQLPBGTPGEDVLQVQRAEAE 1074  
 943 GYQKRLDATALERSNPL-----AAEFKLDLTTRKMHEGPLTWIRISKDTLDQLVILE 998  
 1075 RPEQAEASGLKKEFDVVLKVDQAEAKTEPPTQGVQ-----TPRESFEKAPQVTE 1127  
 999 DLVVLQOEERLLLKCHSKTAVGSSDSKQTFSPVLKNAVLRVATDKRAFLPICSE 1058  
 1128 SI-----ESSELYITTCQAEFLTAGVKSQEMWQEAIPDSV---ETPTDSE-----TDG 1172  
 1059 LGPQIYELVALTSSDKNIMMELLEAVQNAKTHPG--AADPIHPS--PPSQGPAYQGS 1115  
 1173 STPVADFAPAGTQKDEIV--EIHENEVHLVPVGRTEAEVAPAKERPPAPSSFFVQEE 1230  
 1116 TSSREINDSEVYHTEKE-----FKKLPL----- 1138  
 1231 TKESKMDT--LEHTDKSVETVSIISTKTEGTQADQYADEKTKDVPFFEGLSIDTG 1289  
 1139 -----EG-----PGPEQR---VQDKOLIAQGEV 1159  
 1290 ITVSREKYTEVAKGEGEAEACKDDALELQSHAKSPSPVEHEMNVQVREKTEAET 1349  
 1160 QEEDF-----ELRTLPRAAPSIDGENRIGITRDP 1189  
 1350 HVNEKLEHETAVTVSEEVSKQLQTVNVPITIDAKEVSSLEGSPPLGQEEAVCTRIQ 1409  
 1190 V-----LLALGPLLMELGLADALEVDENLR--HLLIMSLPCHYTKQAAGEPE 1237  
 1410 VQSSASTTLTAAAEKEVKVLEGTANILETGETLEBPAGAHVL-----EKKSSSEKN 1459  
 1238 DDLTPTPSVSITSHPMWGPSPQAPITSDSTRLARPEGSQPEGEDVAVASLAILPPTR 1297  
 1460 EDPAHNPEDAV---PTGPDQAKSTPIYASATTKKGLSDBLEK- 1503  
 1298 SSGVWDSPELDRNPAEAASTEPAASYKVRKVSLLPGGVGAAKVAGSNAIPD----- 1351  
 1504 TSLKWSDEVDQVACQEVKVSVAIE-----DLEPENGILIEETKSSKLWNINIGTAV 1556  
 1352 -----SGQSESELSEFVEGGAATGNCIFYVSMAPRPLDSSTPTPTGSPSCSHLPAMPT 1406  
 1557 DQFWRTETATEMLTSELQTA---HYIKADSDAAGQETKEEPEEPASQADETPITSA 1612  
 1407 EPOPYRGVGGCCSILVRDQV---VIFPTIQLTILKRLMDLAEHLELSIGE 1461  
 1613 KEES-ESTAVQOASDISDKMSSEAKTMYEVGSGTYNQDLAEVYLPSSE--EGGG 1667  
 1462 SSGGTFVGSFHTEAARWTDVSLSPPAKEALASQNGQEGSC----- 1505  
 1668 AGTSVPEDDDHALLAEIKESLVEPKEDGDDVDVDPENQNSALADTASGGLTKESPD 1727  
 1506 -----PEGSDALED-----SATDPAVSP 1525  
 1728 TNGPKQKEKEDAEVELOEGKGVHSESDKAITP 1759

RESULT 15  
 US-08-994-570-5  
 ; Sequence 5, Application US/08994570  
 ; Patent No. 6090929  
 ; GENERAL INFORMATION:

```

: APPLICANT: Scott, John D.
: APPLICANT: Naurel, Brian J.
: APPLICANT: Klauck, Theresa M.
: TITLE OF INVENTION: Protein Binding Domains of Gravin
: NUMBER OF SEQUENCES: 24
: CORRESPONDENCE ADDRESS:
: ADDRESS: Marshall, O'Toole, Gerstein, Murray & Borun
: STREET: 6300 Sears Tower/233 South Wacker Drive
: CITY: Chicago
: STATE: Illinois
: COUNTRY: United States of America
: ZIP: 60606-6402
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/994,570
: FILING DATE:
:
: CLASSIFICATION:
: ATTORNEY/AGENT INFORMATION:
: NAME: No. 6090929and, Greta E.
: REGISTRATION NUMBER: 35,302
: REFERENCE/DOCKET NUMBER: 27866/33451
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 312-474-6300
: TELEFAX: 312-474-0448
: TELEX: 25-3856
:
: INFORMATION FOR SEQ. ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1780 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
: US-08-994-570-5

```

Query Match 2.6%; Score 203.5; DB 3; Length 1780;  
 Best Local Similarity 17.0%; Pred. No. 2.7e-06;  
 Matches 312; Conservative 259; Mismatches 704; Indels 557; Gaps 73;

```

QY 29 SSLSISGDSPTSPSHHROPDSTETAGLVORCVIIQKDHGFGFTVSGDRIVLYOS 88
DB 150 SSESNIIEIL-----QPTESQANDIGFKVFKV-----GKFTYKAKTKTEKPT 194
QY 89 VR-----PGAMAKAGVKEGD--RIIKVNGTMYTNSSHLEVYKL 125
DB 195 VOLLYTKDEGEAGACGDHODPSLAGEAASKSEPKOSTEKREPTLKRQSHAEI--- 251
QY 126 IKSQATVALLTGLSSPPS-----VGVSGLDQNSVAGVLRVNPRIIPPPPPPL 174
DB 252 -----SPPAESSGAVECKEKEGKEKPEPSKSA-----ESPT 284
QY 175 PPOHITGPKPLDDPEVOKHATQ--ILM-----NMLRQEEELDILIPPCGETSOTPCGR 228
DB 285 SEPTSTSTG-----STFKKFTTQGMAGMKKTSFKRKPEDEV-----ASEKK 326
QY 229 LSVDSQAGSLDSTGEREPSSIESLMNRNSVLSDPGSDSPQTSPIVLARVAQHRRQGS 288
DB 327 KRQEPKAVTDEDGKAE-----VASEKLTGASEQAHPOEPASAEHPRLSAE 372
QY 289 -DAALLPLNHQ-----GIDQSPKPLITGPEEDYDPCYNNESDITFQ----- 329
DB 373 YEKVELPSEEQVSGSGPSEEPAPL--ATEVFDEKIEVHOEEVVAEYHNVSTVEERTEE 429
QY 330 ---DLEKLR-SHPAYLVFLRYLILSOADPGPLFYLCSEVYOOTNPKDSRLSKDQWINTF 385
DB 430 QATEVETAGSVPAEELVGDAPQAPAKELVKLETCVSGEDPTQAGADLSPD----- 484
QY 386 LKKNAPLRVYKIPMLOAEIDL-----RLRNNEPDRNVLCQAEAVMLLEIOQINDYRSR 440
DB 485 -EK-----VLSKPPEGVYSEVEMLSQERMKVQSGPLKKLFTSTGLKLSKKQ-----KGR 536

```

```

QY 441 TLIGSLYGENDLLGLDGPLRERQMAE-----KQIALGLDILSYEDRSAPMDF 491
DB 537 G-GGDESGEHNTVPADSPDSQOEQKGSASSPEEPETLEKELAEVQD----- 588
QY 492 AVNTFMSHAGIRLRBSRSSCTAKTOSAPDKQWLPFFFKTKQSSNSKKEKDALDKR 551
DB 589 -----GEAEKATSDGEKKRGGVTPWASF-----KKMVPKRVRRPSSDKR 631
QY 552 NPILRYIGKPKSSOSIKKGNVNIITQHFENSHQYVVPFGTORLSTGSFPEDLSDSS 611
DB 632 DE-----LDKYKATLSTSTESTASQEMBEKGSVEEPKRPPEPKRVSTVSWELLICVSS 687
QY 612 RSEIRLGRSGSLKGREEMKRS--RKAENVPRPSDVMDMAAPARLHOSASSAS----- 665
DB 688 KKRARRRRSSDEEGCPKAMGQHQKADENAGKDK-ETGTGDIILAGSOEHPGCGSSSPEDQ 746
QY 666 -----SIST-----KSLNPPTPTPKGRGRISBPNGICFDVILLPHLEDDLG---- 710
DB 747 GSPTGEGSVTWSFRRVLPTRKKSKELENS-----EDSLGSGV 788
QY 711 --QLSDLEPEPEYQNMQHTVGKDVVANLTQREIDROE--VINELFYTEASH--LFTLRVL 764
DB 789 EHSTPTEPEGKE-ESNV-SIKKTIPIRRKKRPDQKQEQAPVEDAGPTGANEDSDVPAYV 846
QY 765 DLIFTQMRKKNIMPPEELARLPNLPLELTHNSCEAMKRIREE----- 810
DB 847 PLSEYDAVERKEMAQO--AKGAEQPE--OKAATVSKELSESQVHMAAANVADCTR 900
QY 811 -GPITIDISDPMARFGDARFARLEQOVAO--FCSYQSALEELIRK-----QKR 857
DB 901 AATLIEERSPSWIS--ASVTEPLEVEAEALITVEYVIAEEPPYTPLEPRK 957
QY 858 ESRFOLFMOEASHP-----QCRRLQRLDIVSEMORLTKYPLLEN 899
DB 958 EARGDTVYSEAEELTPAAYTAETAGPLGSEBGETEASAEETTEMVAVSGLDSDPTTEE 1017
QY 900 I--IKHTEGTSHEKLCARDOCREILKFNVAVR-----QENHRIE 942
DB 1018 APFVQVEEGGVPIEEOER--RTQEVLOAVAEKVEESQLPRTGPGPEVDLPVQVAAE 1074
QY 943 GYKRLDATALERASNP-----AAEFKSLDLTRKKNHSGPLTWRSKDTLDLQVLLLE 998
DB 1075 REPEQAEASGLKKEIDVYLVKDAQEAKEPTFTQGVQVQ-----TTSESEKAPQYTE 1127
QY 999 DLVYLLQROEBRLLKCHKSTAVGSSDKQTPSPVLKLAVALIRSVATDKRAFTICTSE 1058
DB 1128 SI-----ESSELYTTQCAETLAGVKSQEMVMEQAIIPDSV--ETPTDSE-----TDG 1172
QY 1059 LGPQYIELVALTSSDKNTIMMELLEBAVONATKKNP--AAPPIHRS--PRGQEPAYQGS 1115
DB 1173 STPVADFDAPGTQKDEIV--ETHEENEVHLVPRGTAEAVPAOKERPPAPSSFEVQEE 1230
QY 1116 TSGRYEINDSEVYHTEKE-----PKRLP----- 1138
DB 1231 TKEQSKMET-LEHTDKESVETYSILSTKEGQEAQYADEKTKOVFFEGLEGSDTQ 1289
QY 1139 -----EG-----PQDPR--VODKOLLNAGEPV 1159
DB 1290 ITVSRKYTEVALKGGTEAECKKDDALELOSHAKSPSPYEREMVYVVEREKTEAEPT 1349
QY 1160 QEED-----ELTLPRAPSLDGENKRTITRDP 1189
DB 1350 HNEEKLHEHTAVTVSEVSKOLLQTVNPIIDGAKEVSSLEGSPPLCGQEAQVCTKQ 1409
QY 1190 V-----LALTGPLMEGLADALADVENLR-----HLIMSLPLGHTVTOAGAEP 1237
DB 1410 VOSSEASFLLTAAEBEKVLGETANILTEGTLERPAHALYV-----EKSSEKN 1459
QY 1238 DDLTPPSVASTSHMPDQSGQAPPTISDSTRLARPEQSGQEGEDVAVSSLAHLPPRTR 1297
DB 1460 EDFAAHPGEDAV--PTGDCQAKSTPVIVSATTKKGLSSDLEGEKT----- 1503

```





**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OK protein - protein search, using sw model

Run on: September 18, 2002, 10:35:58 ; Search time 45.49 Seconds  
(without alignments)  
3728.505 Million cell updates/sec

Title: US-09-695-795-4  
Perfect score: 7883  
Sequence: 1 MSIRLPHSIDRSASKKQSHL.....EGSDIALEDSATDTAVSPGP 1527

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*  
2: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*  
3: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*  
4: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*  
5: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*  
6: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*  
7: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*  
8: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*  
9: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*  
10: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*  
11: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*  
12: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*  
13: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*  
14: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*  
15: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*  
16: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*  
17: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*  
18: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*  
19: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*  
20: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT:\*  
21: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*  
22: /SIDSL/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7883	100.0	1527	22	AAU01184
2	6455.5	81.9	1520	20	AAV41010
3	1506.5	19.1	1461	19	AAW64468
4	1506.5	19.1	1461	22	AAW64468
5	1216.5	15.4	919	20	AAV41011
6	1196.5	15.2	912	19	AAW42103
7	1194.5	15.2	912	20	AAV41009
8	1100.5	14.0	846	22	AAW12310
9	1100.5	14.0	846	22	AAW79614
10	1100.5	14.0	846	22	AAW41561
11	1068	13.5	848	22	ABG18434

12	945.5	12.0	2559	20	AAV41012	Amino acid sequenc
13	945	12.0	203	22	AAU21680	Novel human neopla
14	936.5	11.9	2559	22	ABW62745	Drosophila melanog
15	892.5	11.3	696	22	AAW39775	Human polypeptide
16	891.5	11.3	696	22	AAW78630	Human protein SEQ
17	700.5	8.9	181	22	AAU17436	Novel signal trans
18	700.5	8.9	181	22	AAU21805	Novel human neopla
19	541	6.9	123	20	AAV41007	Human KIAA380 prot
20	539.5	6.8	169	22	AAW1141	Human immune/haema
21	531	6.7	1050	22	ABW44551	Human wound healin
22	517.5	6.6	1429	20	AAW3941	Human btx protein.
23	517.5	6.6	1429	22	ABW5537	Novel human diagno
24	513.5	6.5	1066	22	ABG15870	Novel human diagno
25	504	6.4	963	22	AAU17312	Novel signal trans
26	433	5.5	96	20	AAV41016	Human KIAA0380 pro
27	428	5.4	418	21	AAW58227	Lung cancer associ
28	410.5	5.2	339	22	AAO10352	Human polypeptide
29	370	4.7	936	22	ABG20576	Novel human diagno
30	346	4.4	1309	22	ABW17862	Drosophila melanog
31	339.5	4.3	1715	21	AAV57449	Mouse Eesl protein
32	332	4.2	235	22	AAW25470	Human protein sequ
33	328	4.2	1697	22	AAW79199	Human protein SEQ
34	322.5	4.1	1683	21	AAV71160	Rat phosphodiester
35	318	4.0	1805	15	AAW60126	Rat nestlin. Ratu
36	318	4.0	1805	15	AAW60126	Rat nestlin protein
37	310.5	3.9	790	22	AAW3045	Human protein sequ
38	301.5	3.8	1658	21	AAV57450	Mouse Eesl2 protel
39	293.5	3.7	1234	22	ABW68510	Drosophila melanog
40	290	3.7	2189	22	AAW79943	Human protein SEQ
41	288	3.7	2161	22	AAW78959	Human protein SEQ
42	288	3.7	2523	22	AAU03503	Human protein kina
43	280.5	3.6	558	20	AAW73379	CSB2 protein sequ
44	279.5	3.5	522	22	AAW5513	Human protein sequ
45	274.5	3.5	1576	22	ABG06283	Novel human diagno

#### ALIGNMENTS

RESULT 1	AAU01184	standard; Protein: 1527 AA.
XX	AAU01184	
AC	AAU01184	
XX	AAU01184	
DT	07-SEP-2001	(first entry)
XX		
DE	Rat glutamate transporter associated protein GTRAP4-48.	
KW	Rat: glutamate transporter associated protein; GTRAP4-48;	
KW	glutamate transporter protein; chloride transport; neurodegeneration;	
KW	cytoskeletal stability; nervous system disorder; schizophrenia;	
KW	spinocerebellar ataxia type I; SCAT; GABA metabolism.	
OS	Rattus sp.	
XX		
PN	W02001.30968-A2.	
XX		
PD	03-MAY-2001.	
XX		
PF	23-OCT-2000; 2000WO-US29431.	
XX		
PR	23-OCT-1999; 99US-0161007.	
PR	22-MAY-2000; 2000US-0206157.	
PA	(UYUO) UNIV JOHNS HOPKINS SCHODI MEDICINE.	
XX		
PI	Rothstein JD, Jackson M, Lin G, Law R, Orlov I;	
XX		
DR	WPI: 2001-300498/31.	
DR	N-PSDB: AAS04274.	
XX		
PT	Novel substantially pure glutamate transporter associated polypeptide	

PT which modulates intracellular glutamate transport, interacts with  
PT glutamate transporter protein and has expression pattern in brain  
XX  
PS Claim 8; Fig 17; 116pp; English.  
XX  
CC The present sequence representing novel rat glutamate transporter  
CC associated protein GTRAP-48 modulates intracellular glutamate  
CC transport, interacts with a glutamate transporter protein and has an  
CC expression pattern in the brain. The GTRAP proteins are also involved  
CC in mediating chloride transport and cytoskeletal stability. The  
CC invention also provides methods for identifying a compound that  
CC modulates a cellular response mediated by a GTRAP protein and a  
CC compound that inhibits the interaction of a glutamate transporter  
CC protein and a GTRAP protein. A compound which modulates the activity  
CC of a glutamate transporter associated polypeptide or interaction with  
CC a glutamate transporter protein is useful for treating a disorder  
CC associated with glutamate transport or chloride flux, preferably a  
CC nervous system disorder such as neurodegeneration, spinocerebellar  
CC ataxia type 1 (SCA1), schizophrenia, epilepsy or a disorder of GABA  
CC metabolism. Such compound could be an antisense oligonucleotide  
CC (AA504278) which can be used for modulating glutamate transport in a  
CC subject.  
CC  
XX  
XX Sequence 1527 AA:

Query Match 100.0%; Score 7883; DB 22; Length 1527;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1527; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSRLRPHSIDRSASKROSHISPIASLWSLSLSDSTPERTSPSHHROPSDSETTAGL 60  
DB 1 MSRLRPHSIDRSASKROSHISPIASLWSLSLSDSTPERTSPSHHROPSDSETTAGL 60  
QY 61 VQRCAVLIQKQDFGFTVSDRIYLVQSVRPGAGAMKAGYKESDRIITKVNTWNTSSHL 120  
DB 61 VQRCAVLIQKQDFGFTVSDRIYLVQSVRPGAGAMKAGYKESDRIITKVNTWNTSSHL 120  
QY 121 EYVKLIKSGAYALTLTGSSPPSVGVSGLOQNPVAGVLRVNTIIPPPPPPLPPPHQHT 180  
DB 121 EYVKLIKSGAYALTLTGSSPPSVGVSGLOQNPVAGVLRVNTIIPPPPPPLPPPHQHT 180  
QY 121 EYVKLIKSGAYALTLTGSSPPSVGVSGLOQNPVAGVLRVNTIIPPPPPPLPPPHQHT 180  
DB 121 EYVKLIKSGAYALTLTGSSPPSVGVSGLOQNPVAGVLRVNTIIPPPPPPLPPPHQHT 180  
QY 181 TGKPLPDPPVOKHAQOILNMNLROEPEELDITLPCGERTSORTCGERLSDVSOEADSL 240  
DB 181 TGKPLPDPPVOKHAQOILNMNLROEPEELDITLPCGERTSORTCGERLSDVSOEADSL 240  
QY 181 TGKPLPDPPVOKHAQOILNMNLROEPEELDITLPCGERTSORTCGERLSDVSOEADSL 240  
DB 181 TGKPLPDPPVOKHAQOILNMNLROEPEELDITLPCGERTSORTCGERLSDVSOEADSL 240  
QY 241 DSGTERPSSISESLMNRNSVLSDPGLSDPQTSFVILARVAQHHRROGSDAALLPLMHQGT 300  
DB 241 DSGTERPSSISESLMNRNSVLSDPGLSDPQTSFVILARVAQHHRROGSDAALLPLMHQGT 300  
QY 241 DSGTERPSSISESLMNRNSVLSDPGLSDPQTSFVILARVAQHHRROGSDAALLPLMHQGT 300  
DB 241 DSGTERPSSISESLMNRNSVLSDPGLSDPQTSFVILARVAQHHRROGSDAALLPLMHQGT 300  
QY 301 DQSPKPLIIPEDDYDPGYNNSSDIIFQDLEKLSHPATLVFLKXIIISQADPGELIFT 360  
DB 301 DQSPKPLIIPEDDYDPGYNNSSDIIFQDLEKLSHPATLVFLKXIIISQADPGELIFT 360  
QY 301 DQSPKPLIIPEDDYDPGYNNSSDIIFQDLEKLSHPATLVFLKXIIISQADPGELIFT 360  
DB 301 DQSPKPLIIPEDDYDPGYNNSSDIIFQDLEKLSHPATLVFLKXIIISQADPGELIFT 360  
QY 361 LCEGVYQQTTPKRSRSLGKINWIFLEKNAPLKVKIPEMIAQEIIDLRLNRNEPRLVCE 420  
DB 361 LCEGVYQQTTPKRSRSLGKINWIFLEKNAPLKVKIPEMIAQEIIDLRLNRNEPRLVCE 420  
QY 361 LCEGVYQQTTPKRSRSLGKINWIFLEKNAPLKVKIPEMIAQEIIDLRLNRNEPRLVCE 420  
DB 361 LCEGVYQQTTPKRSRSLGKINWIFLEKNAPLKVKIPEMIAQEIIDLRLNRNEPRLVCE 420  
QY 421 AOEAVMLEIOEJOINDYRSKRTGLGSLYGENDLLGLDGPRLRERQAEKOLAALGDTLSK 480  
DB 421 AOEAVMLEIOEJOINDYRSKRTGLGSLYGENDLLGLDGPRLRERQAEKOLAALGDTLSK 480  
QY 421 AOEAVMLEIOEJOINDYRSKRTGLGSLYGENDLLGLDGPRLRERQAEKOLAALGDTLSK 480  
DB 421 AOEAVMLEIOEJOINDYRSKRTGLGSLYGENDLLGLDGPRLRERQAEKOLAALGDTLSK 480  
QY 481 YEEDRRAPMDFAVNTFMNSAGITRLRERSRCTAEKQASADKKMLPFPFKTKKQSSNSK 540  
DB 481 YEEDRRAPMDFAVNTFMNSAGITRLRERSRCTAEKQASADKKMLPFPFKTKKQSSNSK 540  
QY 481 YEEDRRAPMDFAVNTFMNSAGITRLRERSRCTAEKQASADKKMLPFPFKTKKQSSNSK 540  
DB 481 YEEDRRAPMDFAVNTFMNSAGITRLRERSRCTAEKQASADKKMLPFPFKTKKQSSNSK 540  
QY 541 KEKDALEDKRRNPILRYIGKPKSSSOSIKPGNRIIIOHENSQOVDPEPTQRLSTGS 600  
DB 541 KEKDALEDKRRNPILRYIGKPKSSSOSIKPGNRIIIOHENSQOVDPEPTQRLSTGS 600  
QY 541 KEKDALEDKRRNPILRYIGKPKSSSOSIKPGNRIIIOHENSQOVDPEPTQRLSTGS 600  
DB 541 KEKDALEDKRRNPILRYIGKPKSSSOSIKPGNRIIIOHENSQOVDPEPTQRLSTGS 600  
QY 601 FPEDLLSSOSSRSEIRLGRSGSLKGREEMKRSKKAENVPRPSDVMDAAEARLHQSA 660  
DB 601 FPEDLLSSOSSRSEIRLGRSGSLKGREEMKRSKKAENVPRPSDVMDAAEARLHQSA 660

QY 661 SSSASSLSTFSLLENPPPTPKMGRSISBPNIICFCTDVLPHLLLEDJGQLSDLEPEPE 720  
DB 661 SSSASSLSTFSLLENPPPTPKMGRSISBPNIICFCTDVLPHLLLEDJGQLSDLEPEPE 720  
QY 721 VQWQWHTVCKDVVANITQREIDROEVINELFYEAASHLRTLRDLITFORMKENLMR 780  
DB 721 VQWQWHTVCKDVVANITQREIDROEVINELFYEAASHLRTLRDLITFORMKENLMR 780  
QY 781 EELARFPNPLPELIEHNSWCENMKLREEGPIIRDISPMLARFDPGAREELOQYAAOF 840  
DB 781 EELARFPNPLPELIEHNSWCENMKLREEGPIIRDISPMLARFDPGAREELOQYAAOF 840  
QY 841 CSYQVALELIRTKKESRFOLEWQEAHPQCRRLQLRDLIVSEMQRITKYPILLENI 900  
DB 841 CSYQVALELIRTKKESRFOLEWQEAHPQCRRLQLRDLIVSEMQRITKYPILLENI 900  
QY 901 IKHTEGTSHEMLCARPOCRILKFNVEAVQNTENRRHREYOKRLATLERSNPL 960  
DB 901 IKHTEGTSHEMLCARPOCRILKFNVEAVQNTENRRHREYOKRLATLERSNPL 960  
QY 961 AAEFKSLDLITTRKMIHEGPIITWRISKDTLDLOVLLLEDLVLLQROEERLLKCHSKTA 1020  
DB 961 AAEFKSLDLITTRKMIHEGPIITWRISKDTLDLOVLLLEDLVLLQROEERLLKCHSKTA 1020  
QY 1021 VGSDDSKQTFSPVLTNAVLINSVATDKRAFTICTSELGPPQIYELVALTSSDKITME 1080  
DB 1021 VGSDDSKQTFSPVLTNAVLINSVATDKRAFTICTSELGPPQIYELVALTSSDKITME 1080  
QY 1081 LLEAVQNAATKHGAPRIPIHPSPGSOEPAYOGSTRREINDSEVYTERKPKKLPG 1140  
DB 1081 LLEAVQNAATKHGAPRIPIHPSPGSOEPAYOGSTRREINDSEVYTERKPKKLPG 1140  
QY 1141 PGEPQVQOKLIAGQEPVQOEDEBELRLPRAPLSDENRGIRTRDPVLLATGPLLM 1200  
DB 1141 PGEPQVQOKLIAGQEPVQOEDEBELRLPRAPLSDENRGIRTRDPVLLATGPLLM 1200  
QY 1201 EGLADALEVENLRHLILMSLLPGHTVKTQANGPEEDDLPTRPSVSTSHPMDSGSP 1260  
DB 1201 EGLADALEVENLRHLILMSLLPGHTVKTQANGPEEDDLPTRPSVSTSHPMDSGSP 1260  
QY 1261 QAPTISDSTRLARPESSQEPEDVAVSSLAHLPPRTSSGWMDSPELDNPAALASTEP 1320  
DB 1261 QAPTISDSTRLARPESSQEPEDVAVSSLAHLPPRTSSGWMDSPELDNPAALASTEP 1320  
QY 1321 AASYKVRKVSLLPGGCGAAGVAGSNALPDSQSSSELSVEBGGAQATGNCPTYSMPAG 1380  
DB 1321 AASYKVRKVSLLPGGCGAAGVAGSNALPDSQSSSELSVEBGGAQATGNCPTYSMPAG 1380  
QY 1381 PLDSSTPEPTGTPSPSOCHSLPMPTEPQYRGVRCGCSLVRPVDVIFHTIQLTLTK 1440  
DB 1381 PLDSSTPEPTGTPSPSOCHSLPMPTEPQYRGVRCGCSLVRPVDVIFHTIQLTLTK 1440  
QY 1441 LHRLEKMEIAHRELKLSLGGESSGGTTPVGSFTEAARTDYSISPAPKALASQONCO 1500  
DB 1441 LHRLEKMEIAHRELKLSLGGESSGGTTPVGSFTEAARTDYSISPAPKALASQONCO 1500  
QY 1501 EGSCPEESDIALEDSATDTAVSPGP 1527  
DB 1501 EGSCPEESDIALEDSATDTAVSPGP 1527

RESULT 2  
AA41010  
ID AA41010 standard; Protein; 1520 AA.  
XX  
AC AA41010;  
XX  
DT 06-DEC-1999 (first entry)  
XX  
DE Amino acid sequence of KIAA380.  
XX  
KW RGS; guanine nucleotide exchange factor; GEF; RGS domain; GAS; cancer;

KW G protein alpha subunit; cell proliferation; growth control; hemostasis;  
 KW morphogenesis; stress fiber formation; integrin-mediated interaction;  
 KW embryonic development; tumor cell growth; cell death; leukocyte homing;  
 KW bone resorption; clot retraction; db1 homology domain; mechanical stress;  
 KW pleckstrin homology domain; KIAA380.

OS Homo sapiens.

PN MO9947557-A2.

XX 23-SEP-1999.

PF 18-MAR-1999: 99WO-US06051.

XX 18-MAR-1998: 98US-0078634.

XX (ONYX-) ONYX PHARM INC.

PI Bollag G, Hart MJ, Roscoe W, Polakis P, Sternweis P, Kozasa T;  
 PI Jiang X;

DR WPI: 1999-571821/48.

DR N-PSDB: AA222282.

PT New isolated RGS-GEF polypeptides, used to develop products for  
 PT modulating, e.g. cell proliferation and integrin-mediated interactions

PS Examples: Fig 12; 75pp; English.

XX The invention relates to isolated RGS-guanine nucleotide exchange factor  
 CC (GEF) polypeptides. The novel isolated RGS-GEF polypeptide consists of an  
 CC RGS domain of a GEF protein and does not comprise a db1 homology (DH)  
 CC domain or a pleckstrin homology (PH) domain. The RGS-GEF polypeptides can  
 CC be used for modulating an activity of a G protein alpha subunit (GAS).  
 CC The products can be used for the regulation of biological pathways in  
 CC which a RGS-GEF polypeptide is involved, particularly pathological  
 CC conditions, e.g. cell proliferation (e.g. cancer), morphogenesis, growth  
 CC control, stress fiber formation, and integrin-mediated interactions, such  
 CC as embryonic development, tumor cell growth and metastasis, programmed  
 CC cell death, hemostasis, leukocyte homing and activation, bone resorption,  
 CC clot retraction, and the response of cells to mechanical stress. The  
 CC products can also be used for detection, diagnosis and production of  
 CC transgenic animals. The present sequence represents the amino acid  
 CC sequence of KIAA380.

XX Sequence 1520 AA:

Query Match 81.9%; Score 6455.5; DB 20; Length 1520;

Best Local Similarity 82.2%; Pred. No. 0;

Matches 1276; Conservative 89; Mismatches 129; Indels 59; Gaps 12;

OY 1 MSIRLPHSIDNSAKSKQSHSSPLASWLSLSSGDSPTSPERTSPSHRQPSDSTSTTACL 60  
 Db 1 msvirpshsidsakskqshssplawlsllsgdsptspertspshrqpsdststtactl 43

OY 61 VORCVTIQKDHGEGFTVSGDRIVYVOSVPRGAAKMGVKGDDITIKVNGTMVNSSHL 120  
 Db 44 vgrcvliqkdngftvgdrlvlgsvrpggaamkagvkgdditlkvnctmynsshll 103

OY 121 EVVKLIKSGAVALLLGSSPPSVSGIQONPVSAGVLRWPIITPPPPPPPLPPPHI 180  
 Db 104 evvklksgavallllgsspsmgisglqdpdpagaprltsvlpssppppplppppri 163

OY 181 TGPKRLDPPEVOKAATQILMMNLROEBELDILPFCGETSORTCESGLSDSGADSGL 240  
 Db 164 tgpkrldppevokaatqilmmnlroebeldilpfcgetsortcesglstdsgadsgl 223

OY 241 DSGTRFSPSISSESLMNRNSVSLDPLDPSPTVLARVAOHRRGGDAALLPLNHGGI 300  
 Db 224 dsgtrfspsisseeslmnrnsvslldpldpsptvlarvaohrrggdaallplnhggi 283

OY 301 DQSPKLLIGDEEDYDPCVFNNESDILFQDLEKLSHPAYLVFLRYLLSQADBPGLLEY 360

Db 284 dqspklligdeedydppyfnnesdillfqdleklshpaylvflryllsqadbppllfy 343

OY 361 LCSEYQOOTNPKDSKSLGKIDWNIFFLEKNAPLRKXIPMLQAEIDLRLNNEPRNVICE 420

Db 344 lcaeayqgaspkdsrslgkldwniffleknaplrkxipmlqaeidrlnneprnvlice 403

OY 421 AOEAVMLEIOEINDYRSKRTLGLSLGENDLLGLDSDPLREBOMAKOLAAIGDITLSK 480

Db 404 ageavmleioeindyrskrtlgslsgendllglstdplrebomakolaagditlsk 463

OY 481 YEEDRSAPMDFAVNTFMSHAGIRLRESRSCTAAKTOSAPRKDWLPFPPTTKKQSSNSK 540

Db 464 yeedrsapmdfavnrtfmsahagirlresrsctaaktosaprkdwlpfppttkkqssnsk 522

OY 541 KEKQALDEKRNPIRLRYTGKPKSSQS-----IKGNVNRNIIQHFENSQYDVE 590

Db 523 kekqaldekrrnpiirlrytgkpkssqs-----ikgnvnrrniqhfnensqyde 582

OY 591 PGTORLSTGSPFPEDLIESDSSRSSEIRLGRSGSLGREMKRSKRAENYPRRSDVMDMA 650

Db 583 pgtorlstgspfpedliesdssrsseirlgrrsgslgreemkrskrakenyprsdvmdma 642

OY 651 AEAARLHQSASSSASLSTRSLLENPTPEPTKMGRRSIESINLFCSTDVILPHILLEDDLG 710

Db 643 aeearlhqsasssaslstsrslletpptekmrrsiesinlfcstdvilphilleddlg 702

OY 711 QLSLDPEPEVONMOHTYKQVYVANLTQRETDROEVINELFVTEASHLRLTVLDLITYQ 770

Db 703 qlsldpepevonnhtykqvvyvanltqretdroevinelfvteashlrltlrvldlityq 762

OY 771 RMREKNLMPREELRLPFLNPDELTEITHNSWCEAKKLEDEEPIRIDPMLAFPDGAR 830

Db 763 rmkenlmpreelrlpflnpdelteithnswceakkledeepiridpmlafpdgar 822

OY 831 EELQOVAQFCSYQVSALELIRTKORKESRQFLQMEASHPOCRRLDLRLVSEMRQL 890

Db 823 eelqovaqfcsyqvallelirtkorkesrqflqmeashpocrrlldrlvsemrql 882

OY 891 TKYPLLENIITKHREGTSEHEKLCRADOCREILKFVNEAVKOTENHRLREGVOKRIDA 950

Db 883 tkypillesitkhregtsehekrcradocrelkfvnvavkotenhrlegvokrida 942

OY 951 TALERASNPILAEFKSLDLTRKMTIHSGPLTMRISKDITLQVLLIEDLVLLQROBER 1010

Db 943 talerasnpilaefksldltrkmtihsgpltmriskditlqvlliedlvllqrodek 1002

OY 1011 LLKCHSTAVGSSDSKQTFSPVILKNAVILRSVATDKRAEFITCTSELAPQIYELVAL 1070

Db 1003 llkchsktavgsdskqtfspvilknavilrsvatdkraeffitctslgppqiyelval 1062

OY 1071 TSSDKNIMELLEFAVONATKHPGAAPRPHSPSGSOPAVOGSTSRVETINSEVYHT 1130

Db 1063 tssdknimellefavonatkhpgaaaprphspsgsopavogstsrvetinsevyht 1122

OY 1131 EKEPKKLPEGGPPEORVODKOLIAOGEPVQ--EDDEELRLPLRAPPLDGENGIRTRD 1188

Db 1123 ekepkklpeggppeorvodkoliaogepvq--eddeelrlplrappldgengirtrd 1182

OY 1189 PVLALATGRLMEGLADALDEVENTLRHLILMSLLPGHTVTKQAAGPEDDLTPPSVVS 1248

Db 1183 pvlalatgrlmegladaldevntlrhlilmsllpghtvtkqaagpeddltppsvvs 1241

OY 1249 ITSHPMDPGSGQART--ISBSTRARREGSOPREGEVAVSSLAHLPRRTSSGVMDSPE 1306

Db 1242 vtsnpmdpgsgqart--isbstrarregsovregeavavsslahlprrtssgvmdspe 1301

OY 1307 LDRNPAALAASTEPAPASVYVVRKVSLLPGGCVGAAKVAGSN---AIPSGGSEELASVEE 1363

Db 1302 ldrnlaeassteaagykvvyrk-----aevagskvypalpesgsgseppreve 1350

OY 1364 GGAQATGNCIFYVSMAPGLDSDSTPEPTGPPSPSOCHSLPAMPTEPQYRGVGGQCS--- 1420



```

Db 959 |cr-maasvkegstkrpdpstpgstpgdgndeedpsklkeeghgstvglqspdralgles 1017
Oy 1107 ---SOEPAYOGSTSSRVEIND-----SEVHTEREPKPLPG-----PCPE 1144
Db 1018 tllskpqshstsgsksevdllfvaergfakegtdcltkeegdyglaipdshlpse 1077
Oy 1145 OR-VOD-----KOLIAOGEPVQED-EHELRLPRAPPSLDSENGIFRRDVLAL 1194
Db 1078 ewaldalrnlgllkqllvqglgleksvgeqwhfpr-----yt-----as 1120
Oy 1195 TGPLMEGLADALAEDEVNLRHLILMSLLPGHTVTOAQAEPEDDLPTPSVYSITSPHW 1254
Db 1121 qgp-----qtdsvlqnsenik---ayhsgeghmpfrtfgtdatcysprtesfa--pr 1170
Oy 1255 DPGSGCAPTISDSTRL-----ARPGS-QPEGEDVAVSSLAHLPPR-TRS 1298
Db 1171 d--svglapqdsqaenllvmthmltptempegegidsgenhfidareahsdenpseg 1228
Oy 1299 SGVMSPELDRN-----PAAEASTEPASVYKVRKVSLLPGGVGAARVA 1344
Db 1229 dgaunkeekdvnlrlisgnyllldgydvqgestdeevs-----sltqpmtdipavest 1283
Oy 1345 GSNA-IPDSGQSESELS-----EVBGAGATGNCFCYVSNPAGPLDSTE----- 1387
Db 1284 hqgghspqnthdgaistpftpeflvgfswamey-scfeiqgspscadsgqimeylnk 1342
Oy 1368 -----PTGPRSPSOCHSLPAMPTEPQPRGVRGQCSLSVARDVDV 1429
Db 1343 eadlhlkegggklrpfakgwldqpsqt-----stqlkvracpggdc-----rllld 1391
Oy 1430 IFHTI-----EQLTKLHRLKDMELAHRELKSLGESSGCTTPVGS 1471
Db 1392 eyrpoltswlqgcgreavt-----edlpynhlglvksqgsgstxsqgs 1438

RESULT 4
AAB90743
ID AAB90743 standard; Protein; 1461 AA.
XX
AC AAB90743;
XX
DT 07-JUN-2001 (first entry)
XX
DE Human CWA20_2 protein sequence SEQ ID 186.
XX
KW Human; secreted protein; nutrient; cytokine modulator; proliferation;
KW differential; immune system modulator; tissue growth; chemotactic;
KW haemostatic; thrombolytic; anti-inflammatory; tumour inhibition;
KW haematopoiesis.
XX
OS Homo sapiens.
XX
PN WO200119988-A1.
XX
PD 22-MAR-2001.
XX
PE 14-SEP-2000; 2000WO-US25135.
XX
PR 17-SEP-1999; 99US-0398829.
XX
PA (GENY ) GENETICS INST INC.
XX
PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;
PI Meiberg D, Treacy M, Bowman MR, Spaulding V, Agostino MJ;
XX
DR WPI; 2001-244801/25.
XX
DR N-PSDB; AAF98482.
XX
PT Isolated nucleic acids encoding polypeptides, useful for modulating
PT e.g. cytokine and cell proliferation/differentiation activity, the
PT immune system and hematopoiesis regulating activity -
XX

```

```

PS Disclosure; Page 505-511; 557pp; English.
XX
CC Human cDNA clones represented in AAF98374 - AAF98489 encode secreted
CC proteins AAB90667 - AAB90750. The cDNA clones are isolated from various
CC tissue types, and may be used in the prevention, treatment and diagnosis
CC of diseases associated with inappropriate protein expression. The
CC polypeptides and nucleic acids may be used as nutrients or to modulate
CC cytokine and cell proliferation/differentiation activity and may also be
CC involved in modulation of the immune system. The cDNA sequences,
CC proteins, their agonists and/or antagonists exhibit haematopoiesis
CC regulating activity; tissue growth activity; activin/inhibin activity;
CC chemotactic/chemokinetic activity; haemostatic and thrombolytic
CC activity; receptor/ligand activity; anti-inflammatory activity;
CC haematopoiesis activity; cadherin/tumour suppressor activity; and/or
CC tumour inhibition activity, included in the invention are probes
CC encoding the secreted proteins.
XX
SQ Sequence 1461 AA;

```

```

Query Match 19.1%; Score 1506.5; DB 22; Length 1461;
Best Local Similarity 29.6%; Pred. No. 4.5e-93;
Matches 459; Conservative 243; Mismatches 494; Indels 357; Gaps 54;

```

```

Oy 161 VNPILPPPPPPPLPPPHITGPKPL--QDEPVOKHATQIILNMNLRQEEELDITLPPCG 218
Db 1 mspilmtspshpsgaasnmerticspvlmgeennvnhqkveillrlnkqegerlql1-----q 56
Oy 219 ETSQTCGRSLVSDSQEADSGIDSGTERF-----DS-----ISESIMN 256
Db 57 edynltqgrllkqlkeqaknhlpqlqeqjlskatsgaqgavvtsrplgdtltvsaead 116
Oy 257 RNSVL-----SDPGLDSPQSPVILARVAQHHRGSDALLRLNHGIDQS 303
Db 117 pgdvlgrtcdssgdasrpsdnadspksgr--kerilylepnexse-tiqdtdcslyvs 173
Oy 304 PK---PLIGPEEDYDGYFNESDII-----FQDLKLSHPAYLVVLRVYLSQAD 353
Db 174 pstriaphlgaedd---dfgtneqngcscqstgellksrphahavflnhvsgfd 229
Oy 354 PEPILFYLCSEVYQOTNPKDSRLGKDIWNIFLEKNAPLRVKYKIPBMLQAEIDL--LRN 410
Db 230 patllcylsdyklhnlknetrrlflefghflinsahlkvsypemsadlekrrpel 289
Oy 411 NEDPRNVCEAQAEMLEIQEINDYRSKRRLGLSLGXGNDLGLDGPRLRERQMAEKQ 470
Db 290 edlnhrylqtmqervhpevgrnlkdfgrkismgl--laeseltkldeardkdritleke 347
Oy 471 LAALGDILSKYE-----EDRSAPMDFAVNTFMSHAGIRLRESRSCTAERTQSAP 520
Db 348 rtcaeqivakleevlmaxgavxwedkstmgvllmymkhlgykvxep-----nle 398
Oy 521 DKDKWLPFPPTTKKOSNSKKEKDALEDKRNPIRLRYGKPKSSQS----- 567
Db 399 hkrgriflplkfq--smkdkkegeekgkrgrfslgpprprphnsaigramejqk 455
Oy 568 -----IKPGVNRNIQHFENSHOYDVPEPQTQLSGSPEDDLEDDSSSEIRLGRS 620
Db 456 arhpkhlistpsvsspepdasaklrsglanegt--dasyjpa-----nmsvasasas 506
Oy 621 GSLKGREEMKRSKRAENVPKRSVDVMDAAEAARLHOSASSASLSTRSLNPTPPF- 679
Db 507 fsqegske-----ndtgsqyvetaspgtdldgtrtlnt-Vdfdfpppld 551
Oy 680 -----TRKMRKSTESPNLGFCTDVIILPHLEDDLGQLSDLEPPEVQN 723
Db 552 qvgeeeceverytehgtkprfk-fdsvalfegsgs-----edeqfe-nleltp--pn 600
Oy 724 WQHTGKDVVNLNQLREIDROEVINELFVTEASHLRTLRVLDLFYQWRKRNLPREEL 783
Db 601 wqqlvsrevllgllkpcelkrqevlnellfyeranvrltkldvqlyqvtrvstregllspel 660

```





```

Db      313 tpgqd-----npgvslhplstsdsvdsrep-gvdtcpgepgdtpppgpcslpep----- 357
Oy      699 VILPHLEDDLGOLS-----DLEPPEVONQNHVYGRKGVANVLTQREIDR 743
Db      358 -lappedstedgetespepgddgepgisgglepe-epppwrelvppdtllslpksqvr 415
Oy      744 QEVINELFVTEASHLRTLRVLDFYQRMKRENLMPREELRLFPNLDELIEHNSWCE- 802
Db      416 gevlsellvteahvrmrlrvhldllygpmadggffrldeqlnifpsidellelvhsifidr 475
Oy      803 AMKRIREGPIIRDISDPMRLARFDPGPARBELOQVAAGFCSYQSAVALELIRTKQKRESRFQ 862
Db      476 lmkrtgeesgylleeigdvllarfdgaegswtqklsficsrqsfaledqkakykrepfc 535
Oy      863 LFMQAEHSHPOCRRLQDLVLYSEMQRLTKYPLLENIKHTEGGTSHHEKLCRARDOCR 922
Db      536 atvgeaesrptcrlqlkdmrlptemqrltkypdllqsglqntleest-ergkveaaeocr 594
Oy      923 ELIKFVNEAVKOTENRHRLEGYOKRLDATALERASNPAAEFKSIDLTRKMIHEGPWTW 982
Db      595 ellhvnqavrdmedllrlkdygrrldtlhltqgsdpmllsefkndltkkklvnegrltw 654
Oy      993 RISKRKITDLOYLLLEDVLYLQREERLLKCHSKTAVGSSDSKQRTSPVLKLANAVLIR 1042
Db      655 rvtckakalevhvllldlllllllqrgderllkshsrlltpcpdqktmlrpyrlrltsamtr 714
Oy      1043 SVATDKRAFFITCSELGPPOIYELVALTSSDKNIMWELLEAQNATKHPGAPPIPIH 1102
Db      715 evatdhkafyylftwd-geaqlyelvaqtsserkwncllletqgslkvpaasrlkprp 773
Oy      1103 SPPGSGEPAYOGSTSSRVEINDESVYHTEKEPKLPBG-----PGPE-----GRVQ 1148
Db      774 spsrlrpepll---ssesengtggaeapadarterllndllpfcrrpgpegqlaatalqvl 830
Oy      1149 D-KOIIAGEVQVEDEDELTLRPAPSLDGEN-RGIRTRDPV-LLALTPPLMEGLAD 1205
Db      831 slkqllls-----teeds-----gaqpprtqdgvyggrtapyvhtqgleenllsleval 879
Oy      1206 AALDVE---NLRLILMSLLPGHTVYTOAAGE 1235
Db      880 rgleeleefcrlrpll--sqllgltlspnlape 911

```

```

RESULT 6
AAW42103
ID AAW42103 standard; Protein: 912 AA.
XX
AC AAW42103;
XX
DT 14-SEP-1998 (first entry)
XX
DE Human p115 Rho-guanine exchange factor.
XX
KW Guanine exchange factor; p115 Rho-GEF; human; GTPase; cancer;
KW metastasis; apoptosis; bone resorption; diagnosis; therapy.
XX
OS Homo sapiens.
XX
PN MO9820127-A1.
XX
PD 14-MAY-1998.
XX
PF 07-OCT-1997; 97MO-US15851.
XX
PR 06-NOV-1996; 96US-0029979.
XX
PA (ONTX-) ONTX PHARM INC.
XX
PI Hart MJ;
XX
DR WPI: 1998-286943/25.
DR N-PSDB: AAV32024.

```

```

XX      p115Rho-guanine nucleotide exchange factor - useful for diagnosis
PT      and treatment of diseases involving GTPase, e.g. cancer, bone
PT      resorption etc.
PS      Claim 4; Fig 1; 66pp; English.
XX
XX      This is the deduced amino acid sequence of novel human p115
CC      p115 Rho-guanine exchange factor (GEF), a polypeptide that has a
CC      specific binding affinity for a guanine nucleotide-depleted state
CC      of G-proteins (in particular RhoA), a guanine nucleotide exchange
CC      activity, an oncogenic transforming activity, and an immunogenic
CC      activity. GEFs modulate cell signalling pathways, both in vivo and
CC      in vitro, by modulating the activity of GTPase. The p115 amino
CC      acid sequence was deduced from a human foetal brain cDNA sequence
CC      (see AAV32024). p115 Rho-GEF and its fragments can be expressed in
CC      recombinant host cells, and used to modulate the activity of Rho or
CC      more generally any GTPase, as well as to screen for (ant)agonists
CC      able to treat diseases or conditions involving GTPase. p115 can be
CC      used to raise antibodies useful for detecting p115 for diagnosis
CC      and research, while fragments of p115 with (ant)agonist activity
CC      are used to treat diseases/conditions where a GTPase is involved,
CC      e.g. cell proliferation (cancer), growth control, morphogenesis,
CC      stress fibre formation, embryonic development, metastases,
CC      apoptosis and bone resorption. Interaction of p115 with Rho
CC      promotes cell proliferation and/or induces cytoskeletal
CC      rearrangements.
XX
SQ      Sequence 912 AA;

```

```

Query Match      15.2%; Score 1196.5; DB 19; Length 912;
Best Local Similarity 34.2%; Pred. No.2.5e-72;
Matches 335; Conservative 167; Mismatches 324; Indels 153; Gaps 32;

```

```

Oy      308 ITGPE-EDYDPGYNN--ESDIITQDLKLSHPAYLVLFRLYLSQADPEPLFYCSE 364
Db      23 llgaeddefeneletnseegnsgfsgleykrrpahalmlqhalqfepgpllcclhad 82
Oy      365 VYQQTNPKDSLSLGDIMNIFLEKNAPLRVKIRPMLAOEID---LRIKNNDEPNVNCIA 421
Db      83 mlgslgpkkaakafldlyhsflektavlrpvpvnafeidrtadlissevqrrtvev 142
Oy      422 QEAVMLEIOEQINDYRSKRRTLGSLYGENDLGLD-----GDPLRROMAEKOLAA 473
Db      143 vsgsqvavgrqldfrkrimgmtr--wegelaglaevwvtrdrasysaerthvaerllmh 200
Oy      474 LGDI--LSKYEDRSAPMDFAVNTFMSHAGIRLRESRSCTAEKTOGAPPKDKWLPPF-- 529
Db      201 leemghstlstdeskaavnaigymthlgyrtksq-----dksgnrfirk 247
Oy      530 -----PKTKKSSNSKKEKDALEDKRRPILRIYIGKPKSSOSIKPGNVRNIIQ 578
Db      248 kvmgnrrsdpprkltkglssi---ldaarwnrgepyvdftrhkaevdaekpg----- 297
Oy      579 HFENSHQDVPEPCTORLSTGSPFEDLLESDSRSSEIRLGRSGSLKGREMKRSKRAENV 638
Db      298 -----atdrkgygvmp-----atd-----rnlgapqgdcpvysllpisl 331
Oy      639 PRPRSDVMDMAAEARLHOSASSASSLSTRLENPTPPFKMGRRSIESPNLFGCTD 698
Db      332 dsprdrepadaplel-----gdsapqgmstleslappedstdeget-esppeg---- 378
Oy      699 VILPHLEDDLGQ--LSDLREP-EYVONQNHVYGRKGVANVLTQREIDROVINYELFTEA 755
Db      379 -----degepyrsglelepeepgywrelvppdtllslpksykyrgevliselllvta 429
Oy      756 SHLRTLRVLDFYQRMKRENLMPREELRLFPNLDELIEHNSWCE-AKKXILREGPIIT 814
Db      430 ahvymrlvlnldlffqmaecelfrlpelqnlfrpsldelelvhsifldrlmkrtggesyyl 489
Oy      815 RDISDPMRLARFDPGPARBELOQVAAGFCSYQSAVALELIRTKQKRESRFQLEMQAEHSHPOC 874

```

```

Db 490 eelgvdllarfdgaegswfqtksrfsrqsfaledqkagqrkdpfcfavgeaesrprc 549
Qy 875 RRLQLDLIVSEMQRLTKYPLLENIKHTEGSTSEHEKLCARADOCREILTKFVNAVKQ 934
Db 550 rrlqldmptemqrltkypdllqsgnteept-erekvelaaceccrellhvnqavd 608
Qy 935 TENRHRLEGYOKRLDTALERASNPPLAEFKSLDTTRKMHGEPJTWIRISKDTLDLOV 994
Db 609 medllrlkdygrldshlrqsgsdpmasefkndltckkrlvhegplrtvrtkckavevny 668
Qy 995 LLELDELVLLOQROEBRLLLKCHSKTAVGSSDSKQTFSPVYLKLANVILRSVATDKRAFFII 1054
Db 669 lllldlllllgqderllllkshsrcltprdgktmlrplrltsamtrevatkafayl 728
Qy 1055 CTSELGPPOIYELVALTSSDKNIMWELLEVAVON-ATKHGAPAPIHPSPOQOPAYO 1113
Db 729 ftwd-geaqlvelvaqteerkmwcalitetagslkvpasprpk-rprpstrelpils 786
Qy 1114 GSTSSRVEINDESVYHTEKEPKKLPG-----PGEQRYOD-----KOLIAQGE 1158
Db 787 sseengr---gretspadarterliledllprcfprpegqiaatalrkvlslkqllf---p 840
Qy 1159 VOEDEDEELRTLRAPPSIDGEMRGIRTRDPVLLALTGPLMEGLAD-----ALEDVE- 1212
Db 841 aeedn-----gagprtdgd--gvp99g9plspartqe-igenllsleemkqleelae 889
Qy 1213 ---NLRHILWSDLPCHTV 1228
Db 890 eefrlrpll--sqlgnsav 906

RESULT 7
AAV41009
ID AAV41009 standard; Protein; 912 AA.
AC AAV41009;
XX
XX 06-DEC-1999 (first entry)
DE Amino acid sequence of p115 Rho GEF.
XX
XX RGS; guanine nucleotide exchange factor; GEF; RGS domain; GAS; cancer;
XX G protein alpha subunit; cell proliferation; growth control; hemostasis;
XX morphogenesis; stress fiber formation; integrin-mediated interaction;
XX embryonic development; tumor cell growth; cell death; leukocyte homing;
XX bone resorption; clot retraction; db1 homology domain; mechanical stress;
XX plectroin homology domain; p115 Rho GEF.
XX
XX Homo sapiens.
XX
XX WO9947557-A2.
XX
XX 23-SEP-1999.
XX
XX 18-MAR-1999; 99WO-US06051.
XX
XX 18-MAR-1998; 98US-0078634.
XX
XX (ONYX-) ONYX PHARM INC.
XX
XX Bolland G, Hart MJ, Roscoe W, Polakis P, Sternweis P, Kozasa T;
XX Jiang X;
XX WPI; 1999-571821/48.
XX N-PSDB; AA22281.
XX
XX New isolated RGS-GEF polypeptides, used to develop products for
XX modulating, e.g. cell proliferation and integrin-mediated interactions
XX
XX Examples; Fig 10; 75pp; English.
XX
XX The invention relates to isolated RGS-guanine nucleotide exchange factor

```

```

CC (GEF) polypeptides. The novel isolated RGS-GEF polypeptide consists of an
CC RGS domain of a GEF protein and does not comprise a db1 homology (DH)
CC domain or a plectroin homology (PH) domain. The RGS-GEF polypeptides can
CC be used for modulating an activity of a G protein alpha subunit (GAS).
CC The products can be used for the regulation of biological pathways in
CC which a RGS-GEF polypeptide is involved, particularly pathological
CC conditions, e.g. cell proliferation (e.g. cancer), morphogenesis, growth
CC control, stress fiber formation, and integrin-mediated interactions, such
CC as embryonic development, tumor cell growth and metastasis, programmed
CC cell death, hemostasis, leukocyte homing and activation, bone resorption,
CC clot retraction, and the response of cells to mechanical stress. The
CC products can also be used for detection, diagnosis and production of
CC transgenic animals. The present sequence represents the amino acid
CC sequence of p115 Rho GEF.
CC
XX
SQ Sequence 912 AA:
Query Match 15.2%; Score 1194.5; DB 20; Length 912;
Best Local Similarity 34.4%; Pred. No. 3,4e-72;
Matches 339; Conservative 165; Mismatches 316; Indels 165; Gaps 33;
Qy 308 IIGPE-EDYDPCGFNN--ESDIIIFODLEKLSHPALVFLVRLISQADPGPLFLGSE 364
Db 23 llygaededefeneletcseeqnsqfqlsqvkrtrpanhmaligvalqfepgpllcclhad 82
Qy 365 VYQQTNPKDSRLSGKDINWIFLEKNAPLRVKIPEMLDAEID--LRLNNEDEPRNVLCBA 421
Db 83 mlgslgpkkaekafldfyhsflektaavlrvpvpnaafeldtraddlsedqgrtfrvge 142
Qy 422 QEAVMLEIOEQINDYRSKRTLGSLXGENDLLGLD-----GDPLEROMAEKQLAA 473
Db 143 vsqgvaavgrqledftrsklmgmtp--wegelaqleaawgrdrasyearvaerlva 200
Qy 474 LGDI--LSKYEEDRSAPMDFAVNTFMSHAGIRLRESRSSCTAEKTOQAPADKDKMLPF-- 529
Db 201 leemqhtlstdeeksaavnaiglymrhlygrtksg-----dkksgrnffik 247
Qy 530 -----PKTKQSSNSKKEKDALDDKRNPLRKTYIGPKSSOSIKGNVNRNTIQ 578
Db 248 kymgnrttsddppktkqylsl---ldaarwnrgepvpadfrhlkaevdaekpg----- 297
Qy 579 HFNESHQYVPEPGQRLSTGSPFEDLSDSSRSIRLGRSGSLKRGEMRKRKAENV 638
Db 298 -----atdrkgygvmp-----sid-----mlgapgqdtpgvslhpsl 331
Qy 639 PRFRSDVMDAAEAARLHQASSSASLSLSTRLENRPFPFKMGGRSIESPNLGFCTD 698
Db 332 dsprdregeadaplel-----gdsppqgmleelappestdegaet-eespep9---- 378
Qy 699 VILPHLEDDLQO--LSDEPEP-EYQNMQHTYGVKQVYANLQREIDRQEVNELFVTQA 755
Db 379 -----degeprgsqlelepeepggrwrelvpdltlshlpsqvkqgevisellvtea 429
Qy 756 SHURTRVLDFLFYOMRKENLMPRELARLPNLPDELHINSWCE-AMKRLREGPII 814
Db 430 ahvtmlrvlhdllfqpmaecifflpeelqnlfrpsidellevmslfdrlmktrrgeegyl 489
Qy 815 RDISPMLARFGPARREELQOVAOFCYSOVALDELIRTKRKESRFQLEFMQEAESHPOC 874
Db 490 eelgvdllarfdgaegswfqtksrfsrqsfaledqkagqrkdpfcfavgeaesrprc 549
Qy 875 RRLQLDLIVSEMQRLTKYPLLENIKHTEGSTSEHEKLCARADOCREILTKFVNAVKQ 934
Db 550 rrlqldmptemqrltkypdllqsgnteept-erekvelaaceccrellhvnqavd 608
Qy 935 TENRHRLEGYOKRLDTALERASNPPLAEFKSLDTTRKMHGEPJTWIRISKDTLDLOV 994
Db 609 medllrlkdygrldshlrqsgsdpmasefkndltckkrlvhegplrtvrtkckavevny 668
Qy 995 LLELDELVLLOQROEBRLLLKCHSKTAVGSSDSKQTFSPVYLKLANVILRSVATDKRAFFII 1054
Db 669 lllldlllllgqderllllkshsrcltprdgktmlrplrltsamtrevatkafayl 728

```

bind to polypeptides of the invention. Although novel, many of the

493 kpgshslstsqksevrdlfvaerqfakeqhtdgtlkevgedyqiaipdshlpvseerwal 55



QY 1304 SPELDNR-----PAAEASTEPAAKYKVVKVSLLPGGCGVGAAYAGSNA- 1348  
 Db 704 keekdvnlrignyllldgypgvsgstdeevas-----slltgmtdipavesthqqh 758  
 QY 1349 IPDSGQSSELS-----EVEGGAOATGNCIFYVMAGPLDSTETPTGPPSPSOCHS 1400  
 Db 759 spgnhtsdgaaspftpeflvqrgwamey-scfeiqspscadsqsg----- 804  
 QY 1401 LPAMPTPEOPRGVGGCCSSLVRRDQVDFIHTTEOLITIKLHRLKMDLAHRELKLSIGG 1460  
 Db 805 -----imeyhlkicadlenhkkveesytllcqrtaag 835  
 QY 1461 ES 1462  
 Db 836 sa 837

RESULT 10  
 AAM41561  
 ID AAM41561 standard; Protein: 846 AA.  
 AC AAM41561;  
 XX  
 DT 22-OCT-2001 (first entry)  
 DE Human polypeptide SEQ ID NO 6492.  
 XX  
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;  
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;  
 KW leukaemia.  
 XX  
 OS Homo sapiens.  
 PN WO200153312-A1.  
 PD 26-JUL-2001.  
 XX  
 PE 26-DEC-2000; 2000WO-US34263.  
 XX  
 PR 21-JAN-2000; 2000US-0488725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-0598042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-AUG-2000; 2000US-0653450.  
 PR 14-SEP-2000; 2000US-0662191.  
 PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AD, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;  
 XX  
 DR WPI: 2001-442253/47.  
 DR N-PSDB; AAI60717.  
 XX  
 PT Novel nucleic acids and polypeptides, useful for treating disorders  
 PT such as central nervous system injuries -  
 XX  
 PS Example 2; SEQ ID NO 6492; 10078bp; English.  
 XX  
 CC The invention relates to human nucleic acids (AA15798-AA161369) and  
 CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,  
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC localised neuropathies and central nervous system diseases, such as  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic

CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilisation of the activities such as: Immune system suppression,  
 CC Activation/inhibition activity, chemotactic/chemokinetic activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukemias and  
 CC C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification.  
 XX  
 SO Sequence 846 AA;

Query Match 14.0%; Score 1100.5; DB 22; Length 846;  
 Best Local Similarity 32.9%; Pred. No. 7.4e-66;  
 Matches 297; Conservative 142; Mismatches 258; Indels 205; Gaps 30;

QY 670 RSLKNPT-PTPKMGRSISPNLIGCTDYLPHLEDDLGQSDLEPEPEVONQHTV 728  
 Db 32 rvtetgtpkpf-----fkfsvafgesqs-----edeqfe-ndletdp--pwwqqlv 75  
 QY 729 GKDVAANLTGREIDROEVINELFTVASHLRTLRVLDLIFYORRKNELMREELARLP 788  
 Db 76 sreylglkpcelkrqevlneltyerahvrlkvldgvlyrsreglspselrkifs 135  
 QY 789 NLPELIEIHNSWCAMKKLRE--EGPITRIDSDPMARFDPAREELOQVAAQCSYQSV 846  
 Db 136 nledlqlghlneqmkvkrknetsevidqgedliltfsgpgeeklkhaattcngpf 195  
 QY 847 ALELRITKORESPOLFMOEAESHPOCRQLQDLIVSEMQRLTKYPLLENTIKHTEG 906  
 Db 196 alemkrsqkdsrftqvtvdaesnprrlqkldlptqmrlltkypdllidnatylew 255  
 QY 907 GTSEHEKLCRARDOCRELKEVNEAVKOTENRHNLEGGQKRLDNTALERAANPLAEKRS 966  
 Db 256 pt-erekvkaadncrqlinyngvkaenkqrlodyqrrldsslkiseyprveelrn 314  
 QY 967 LDLTRKMINHGPRLTWRISKQRTDLYLLEDDLVLLROEROERLLKCHSKRTAVGSSDS 1026  
 Db 315 ldlckrmihgprlwkvrncktdllylllledlilvlkqddrlvlrchskllastads 374  
 QY 1027 KOTSPVLKLNANLIRSATTKRAFFITCTSELPPOLYELVALTSSKKNIMLEDEAV 1086  
 Db 375 khtsprvklstvlvrgvatnkalfvismedng-aqlyelvaqvtsekwqlllcr-m 432  
 QY 1087 QNARKHGAAPLPIHPSPG-----s 1107  
 Db 433 aasvkegstckpdlpqscpgegdndeepsklkeeqngisvtqjgsprdtglsecliss 492  
 QY 1108 QEPAYOGSTSSRVEIND-----SEVYHTEKEPKKLEPG-----PGPEOR-VQ 1148  
 Db 493 kpgshlatsgsksevrldfvaerqfakqhtdgtlkevgedylqalpshlpvseerwa1 552  
 QY 1149 D-----KOLIAQGEVQGED--BEELRTLPRAPPSLDGENRGITRPDVLALTGRL 1199  
 Db 553 dalrnlglkqlvqqlgltkvsqvedwqhfr-----yrt-----asqgp-- 593  
 QY 1200 MEGGLADALEVDENRHLILMSLPGHTVKTQAGPEPDDLTTPRPSVYSIHMPDPESP 1259  
 Db 594 ---qtdsvlqnsenlk---ayhsgeglmpfrtgytdatcysprstestfa--prd--sv 643  
 QY 1260 GQAPRTISDTRL-----ARPEGS-OPEGEDVAVSSLAHLPPR-TSSSGVWD 1303  
 Db 644 glapqdsqasnllymdhmtprempmepegldsgchffidareahsdempsegdgaqn 703  
 QY 1304 SPELDNR-----PAAEASTEPAAKYKVVKVSLLPGGCGVGAAYAGSNA- 1348  
 Db 704 keekdvnlrignyllldgypgvsgstdeevas-----slltgmtdipavesthqqh 758  
 QY 1349 IPDSGQSSELS-----EVEGGAOATGNCIFYVMAGPLDSTETPTGPPSPSOCHS 1400  
 Db 759 spgnhtsdgaaspftpeflvqrgwamey-scfeiqspscadsqsg----- 804  
 QY 1401 LPAMPTPEOPRGVGGCCSSLVRRDQVDFIHTTEOLITIKLHRLKMDLAHRELKLSIGG 1460

DB 805 -----lmeYlHkIeadleHlkKveesYtlIcgrIag 835

QY 1461 ES 1462

DB 836 sa 837

RESULT 11

ABG18434

ID ABG18434 standard; Protein; 848 AA.

AC ABG18434;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #18425.

XX

KM Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder.

XX

OS Homo sapiens.

XX

PN W0200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

XX

N-PSDB; AAS82621.

PT

PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity

XX

PS Claim 20; SEQ ID No 48793; 103pp; English.

XX

XX The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers, and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX

SO Sequence 848 AA:

Query Match 13.5%; Score 1068; DB 22; Length 848;

Best Local Similarity 32.3%; Pred. No. 1.2e-63;

Matches 281; Conservative 137; Mismatches 239; Indels 212; Gaps 25;

QY 713 SDEPPEVOMQHTGVKDVANLTQREIDROEVNELFVTEASHRLRTLRVLDLIFYORM 772

DB 54 ndletdp--pnwqqlvsrevllglkpoelkgevinelfyterahvrtlkvidqfyqrv 111

QY 773 RKENLMPREELARLFNLPPELLIHNHSCVAMKLE--EGPIIRDISDPMARFGPAR 830

DB 112 sregllspselelrfkifsnldqlqihglnegmkavkrnetsvidqldgedllwfsagpe 171

QY 831 EELQOVAACFSYQVALLEIRTKORKESEFOLFMEQAEASHPCCRRLQRLDLVSEMR 890

DB 172 ekikhaatfcnsqfalenihsrqqkdsrfgfvdgaesnpccrrllqldilpvtqmqrl 231

QY 891 TKYPLDLENIKHTEGTSHEKLCRAPDQCRREILFVNEAVKQTENRHL--EGYOKRLD 949

DB 232 tkypillldnlakylewpt-erekvkaadncrgllnfvgavkeankqgleedygrld 290

QY 950 ATALERASNPAAEFSSDLITRKMHHEGPIWTRISKDTLIDQVLLLELVVLLROEE 1009

DB 291 tselkiseypnverlrnmldtkrkmlhegplvkvkrndrktldlyllledllvllqkqdd 350

QY 1010 RLILKCHSKTAVGSSDSKOTFSPVLKLNALVLSVATDKRAFFICTSELGPQIYELVA 1069

DB 351 rlvtrchsklstaadskntfsviklsvlyvqvadnkalfvismdng-ajlyelva 409

QY 1070 LTVSSDKNIMWELLEAVONATKHPGAAPIPHPSPGSEPAYOGSTSKREIINDESVYH 1129

DB 410 qtvsekltwqdlfcr--maasvkqscckpripsttpgqd-----nd----- 450

QY 1130 TEKEPKKLPGE-----PGPEQRVODKQLAAGEP-----VQEDDEELRTLPAPPS 1176

DB 451 -eepesklkeeghglsvtlqspdrldglesclsskpsghsistgsksevirllfaerq 509

QY 1177 LD-----GENRGIRTRDPVLLALTGPLMEGLADALEDVENLRHLIT----- 1219

DB 510 fakeqhtdgtlkevgedyalrpsdnl-----pvskerraladelrnlglkqlvhnqgl 564

QY 1220 -----WSLPGHTVTKQAAGEPDDLTPTPSVVSITSHPMWDPGPGCAPTISDSTR 1271

DB 565 teksalenwqhfprryrtafg---pqtd-----sviq----- 593

QY 1272 AREPGSQPEGEDVAV--SSLAHLPRTRSSGVWDSPELDNRNPAEAASTEPAASYVVRK 1329

DB 594 -----nsenlkayhsgeqlmfrtgcgdl-----atcysprstse----- 628

QY 1330 VSLPFGGVGAAKVAGSNAPDSGQS-----ESELSEVGEAGAOATGNCFYVS 1376

DB 629 -sfaprdsvyla-----pqdsqasnllvndhmimtpempegglddsgehffda 679

QY 1377 MPAGPLDSSTPTPTGPPSPSQCHSLPAMPTEPOPYRGVNGGCCSLVRRDVDVIFHTIQ 1436

DB 680 reahsdensegdy-----avnskeadv----- 702

QY 1437 LTIKHLKLMELAHLELKSIGSESSGTTVPVSGFHTFAARMTDVSLSPAKALASPS 1496

DB 703 -----nlrtisgnyllld-----gydprgeestdeavasslsqpmtylpaves 746

QY 1497 QNGQEGSCPEGSIDALDSATDPAVSP 1525

DB 747 -----plqgatslrnthsdgelsp 766

RESULT 12

AAV41012

ID AAV41012 standard; Protein; 2559 AA.

XX

AC AAV41012;

XX

DT 06-DEC-1999 (first entry)

XX

DE Amino acid sequence of DRHOGEPZ.

XX

KW RGS; guanine nucleotide exchange factor; GEF; RGS domain; GAs; cancer;  
 KW G protein alpha subunit; cell proliferation; growth control; hemostasis;  
 KW morphogenesis; stress fiber formation; integrin-mediated interaction;  
 KW embryonic development; tumor cell growth; cell death; leukocyte homing;  
 KW bone resorption; clot retraction; dbi homology domain; mechanical stress;  
 KW pleckstrin homology domain; DrhGEF2.  
 XX  
 OS Drosophila sp.  
 PN MO9947557-A2.  
 XX  
 PD 23-SEP-1999.  
 XX  
 PE 18-MAR-1999; 99WO-US06051.  
 XX  
 PR 18-MAR-1998; 98US-0078634.  
 XX  
 PA (ONYX-) ONYX PHARM INC.  
 PI Bollag G, Hart MJ, Roscoe W, Polakis P, Sternweis P, Kozasa T;  
 PI Jiang X;  
 XX  
 DR WPI: 1999-571821/48.  
 DR N-PSDB: AAZ22284.  
 XX  
 PT New isolated RGS-GEF polypeptides, used to develop products for  
 PT modulating, e.g. cell proliferation and integrin-mediated interactions  
 PT  
 PS Examples: Fig 16; 75pp; English.  
 XX  
 CC The invention relates to isolated RGS-guanine nucleotide exchange factor  
 CC (GEF) polypeptides. The novel isolated RGS-GEF polypeptide consists of an  
 CC RGS domain of a GEF protein and does not comprise a dbi homology (DH)  
 CC domain or a pleckstrin homology (PH) domain. The RGS-GEF polypeptides can  
 CC be used for modulating an activity of a G protein alpha subunit (GAs).  
 CC The products can be used for the regulation of biological pathways in  
 CC which a RGS-GEF polypeptide is involved, particularly pathological  
 CC conditions, e.g. cell proliferation (e.g. cancer), morphogenesis, growth  
 CC control, stress fiber formation, and integrin-mediated interactions, such  
 CC as embryonic development, tumor cell growth and metastasis, programmed  
 CC cell death, hemostasis, leukocyte homing and activation, bone resorption,  
 CC clot retraction, and the response of cells to mechanical stress. The  
 CC products can also be used for detection, diagnosis and production of  
 CC transgenic animals. The present sequence represents the amino acid  
 CC sequence of DrhGEF2.  
 XX  
 SO Sequence 2559 AA;  
 Query Match 12.0%; Score 945.5; DB 20; Length 2559;  
 Best Local Similarity 20.9%; Pred. No. 1.2e-54;  
 Matches 482; Conservative 280; Mismatches 650; Indels 895; Gaps 82;

OY 171 ----- 170  
 Db 504 lslgtrkntekdltspsflltdllqgqrmsbqasmsgmhstcptscqfllphq 563  
 OY 171 -----PPLP---PQ-----HTTG 182  
 Db 564 qghrfketgtskgnkflslrsllleedvppplpqrrppqllndllknasppgsnlva 623  
 OY 183 P-----KPLQDPEVQKHATQILNMN 202  
 Db 624 pvsdlratatpqlnrsgqqlprstlndspnsakrskiklkaltdp---kmsqmlilgm 680  
 OY 203 ----- 202  
 Db 681 esasaagaagslewdgppplpprlpgmmtemdsrgscqnlagnsvgtfnylvsct 740  
 OY 203 -----LROEDELQDIL-----PPCGETSQR----- 223  
 Db 741 tavqndlnlafplsqrplvqqlqyqgqghqmsggatgalqtpnlqknkhrvys 800  
 OY 224 -----TCGRSLSDVDSQEADSGLDGSTERFESISLSM-----NRNS 259  
 Db 801 spdmhprhprlrlklttsqswelvedgess-ppqtpilpylssshmtvledpneumrga 859  
 OY 260 VLSDPG--LDSPO-----TSPVILARVACH-HRROGSDAALLPLNHQIDOSPRLTI 309  
 Db 860 aaagpgvfieshqftfmgagaaspilslshshmbaagnd-----tqkeils 906  
 OY 310 GPEDDYD-----PGYFNESDIIIFQDLLEKLSHPAYLVFLRYLSQADPGPLLEYL 361  
 Db 907 medensdidepfidengfnlrltrlleae-----nvflaifllynsadppllfyl 960  
 OY 362 CSEFYQGNPKDSSLGCDINIFLEKNAPLR-VKIPMLQAEIDLRLRNEDP-----RN 416  
 Db 961 ltefyegtskdmckwayelnsstlvpaplsyqdeslarevduvqlqeydveillrt 1020  
 OY 417 VLCAQEVMEIIOEINDYRSKRTLGSLGSEYEND--LLGLDDPLEROMAKOL--- 471  
 Db 1021 vflsrkrakallseqleqfegkrtagqgtlygptddklaekdklire-qldkylmpn 1079  
 OY 472 -----AALGDIILSKYEEDRSAPMDFA--VNTFMSH----- 499  
 Db 1080 lhaliedengspedvtrkvalcalstvllyrlftrppssilvervhfvarfskari 1139  
 OY 500 -----AGIR-----LRE-----SRSCTAERKQS 518  
 Db 1140 mgknrkmnvgrhplvlrlyevthcnhcgllwvsgpgyhtcdklnlhrqcskvvdes 1199  
 OY 519 APDKDKMLPFPFKTKRKOSNSKKEK-----DALEDKR----- 551  
 Db 1200 cpgp-----lpgakrlahnakiskfmgklrptscdvngnekrtrgdeidvelltprdq 1253  
 OY 552 -----NPLIRYIGKPKSSQSIRKGNVRNIIQHFENSHQDYDEPG----- 592  
 Db 1254 asivrgpsdrpdanlsirngtsntsgl---ntdlqsgfngscandsinp9gagc 1310  
 OY 593 TORLST-----GSFPR-----DLLESSSR-----SEIRLGRS 620  
 Db 1311 nmdlstsvasstptsgsvaaaglsafaelnaidvckearreyrsqphshsapsvnr 1370  
 OY 621 GSLGREEMKRSRAEN-----VPRPSDVMDAAEARLHQASSSASSTLS----- 668  
 Db 1371 esykerlsnktrnrstkspslsstip-ndeqldlqslnatlygssnsslsasggesp 1429  
 OY 669 TRSLENPTRP-----FTPKMGRSIESPNL-----GCTDYVILPHLEDDIGOLS-- 713  
 Db 1430 stmehfagaagaagvypnpqllhllqhaqyqgqsfqaglaagaasas 1489  
 OY 714 -----DLPEPEYQ--NMQHVGVGDVYANLQRELDROEVINELFVT 753  
 Db 1490 nssfwaghpilparwrlteseddvneadwsmvaeavlalldaekkrqellineiyqt 1549

QY 754 EASHLTATRVLDLIFQRRKKNLMPRELARLF-PNLPELITIHNSWCAMKKLR-EEG 811  
 Db 1550 ernhvtrtkldrlffrlpyesglisqdhlllfpallslreihqafesglxqrlteln 1609  
 QY 812 PIRIDSDPLARPDGPARRELOOVAOPCYSVALERLRTQKRESRQLEMOAESH 871  
 Db 1610 hvvnttgdlladmfddqsgyvvlcefaqfcarqialaekrnkdemlqklkksesh 1669  
 QY 872 POCRLQLRDLIVSEMOURLTKYPLLENIKNH-EGGSTSEHKLCARADOCREILKE 927  
 Db 1670 kacrrtelkdlptvqlrtkyppllfenlykvtvrlpentteaaiqraveskrlve 1729  
 QY 928 VNEAVVQTEKRRRLLEGYQKRLDATALERASNPILAEFKSDLTRKMIHEGPIITWISKD 987  
 Db 1730 vngavvtaedanklqnqtkldrsysyk-----eefkklldlqhlhngnlt-ikkn 1781  
 QY 988 KTLIDLVLLEDVLVLQOEERLLK-CHSKTAVGSSDSKQTFSPVLKLN-VLIRSVN 1045  
 Db 1782 psvqlnglllfemvllltqddkyllknhtplsi-----tnkvspimsidadtllirgea 1837  
 QY 1046 TDKRAFFICTSELGPPIYELVALTSDBKNIMWELLEAVONATRH----- 1092  
 Db 1838 adksnffll---kmtksqmlrlapsssecktwfkhsdvaarqsknrksnhdtsl 1894  
 QY 1093 --PGAPPIRHSPPSQOE-----PAYGSTS 1118  
 Db 1895 sdpaalaip-hsnukeslstdltvgplaataltlttjplamjlatvtpatnnsnv 1953  
 QY 1119 -----RVEINDEYVHTEKEPKKLPEEGPQORVODKOLI-AGEPYO-----EED 1163  
 Db 1954 sltgvqlrpprdataseadaqvyntpkprsgnevrnctmslstgeplqkyaanltean 2013  
 QY 1164 EEEELR---TLPPRPSLDGENR-----GIRTRDPVLLAL-----TGPILMEG-LA 1204  
 Db 2014 dvlrlhsgstresvrgstgeenstygmgyskrdasivcsnnsntrtllmgsplv 2073  
 QY 1205 D-AALDEVDENLRLLIIMSLPGHTVKTQAAAGEBED-----DLTPRP-----SVVS 1249  
 Db 2074 dptatvstispahtaepvltgkrlrlldaslrndllekqkicldfrlprvehydaqivdi 2133  
 QY 1250 TSHPMDPGSPGQAPRTISDSTRLARPEGSQ-----PEGEDVAVSSL-AH- 1291  
 Db 2134 ammp-----eapksadadalaaydqitlctkmleymhvtgpegvsavstavgchne 2186  
 QY 1292 -----LPPTRSSGVWDSPEPLDRNPAAEASTEPASIKYKRVSVLLPGCGVGAAY 1343  
 Db 2187 keklrkkvapsssfssppplppnrxhaqagqilpps---rlmpklqlldidev----- 2238  
 QY 1344 AGSNALPDGQSSELSFEVGGAGQATGNCFYVGMPPAGPLDSSREPGTPPSPGCHSLPA 1403  
 Db 2239 ----alheddadyceidelt-----lpaipskpherpt-tplap----- 2272  
 QY 1404 WPTPEQRYRGVRCGQCSLVRDVIIFHTI-----EQLTIKLRH- 1444  
 Db 2273 fhteptsgsv-----idaaskrgstadvpeglllegepldgkeltkgednevktvpsdkls 2228  
 QY 1445 ----KMEIAHRLKLSLGGESSGGTTPVGSFTEAARMTDVSLSPAKEALASDSQNG 1500  
 Db 2329 escneerqcvaeadltkevdpdltskneaaasvdelpsqsrrelktaenasksvadkchedne 2388  
 QY 1501 EOGSCPEEGSDIALEDSATIDTAVSPG 1527  
 Db 2389 e---ltiegv-astvdsstqstpsesp 2411

RESULT 13

AAU21680 standard: Protein: 203 AA.

AAU21680;

04-DEC-2001 (first entry)

XX

DE Novel human neoplastic disease associated polypeptide #113.  
 XX Human: neoplastic disease associated polypeptide; cancer;  
 KW hyperproliferative disorder; neural disorder; immune system disorder;  
 KW muscular disorder; reproductive disorder; gastrointestinal disorder;  
 KW pulmonary disorder; cardiovascular disorder; renal disorder;  
 KW neuroprotective; cytostatic; anti inflammatory; vasotropic.  
 XX  
 OS Homo sapiens.  
 PN WO200155163-A1.  
 PD 02-AUG-2001.  
 XX  
 PF 17-JAN-2001; 2001WO-US01358.  
 XX  
 PR 31-JAN-2000; 2000US-0179065.  
 PR 04-FEB-2000; 2000US-0180628.  
 PR 24-FEB-2000; 2000US-0184664.  
 PR 02-MAR-2000; 2000US-0186350.  
 PR 16-MAR-2000; 2000US-0189874.  
 PR 17-MAR-2000; 2000US-0190076.  
 PR 18-APR-2000; 2000US-0198123.  
 PR 19-MAY-2000; 2000US-0205515.  
 PR 07-JUN-2000; 2000US-0209467.  
 PR 28-JUN-2000; 2000US-0214886.  
 PR 30-JUN-2000; 2000US-0215135.  
 PR 07-JUL-2000; 2000US-0216647.  
 PR 07-JUL-2000; 2000US-0216880.  
 PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225267.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226682.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.



PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 25-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235634.  
 PR 27-SEP-2000; 2000US-0235636.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.  
 PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 01-NOV-2000; 2000US-0246474.  
 PR 08-NOV-2000; 2000US-0246475.  
 PR 08-NOV-2000; 2000US-0246476.  
 PR 08-NOV-2000; 2000US-0246477.  
 PR 08-NOV-2000; 2000US-0246478.  
 PR 08-NOV-2000; 2000US-0246523.  
 PR 08-NOV-2000; 2000US-0246524.  
 PR 08-NOV-2000; 2000US-0246525.  
 PR 08-NOV-2000; 2000US-0246526.  
 PR 08-NOV-2000; 2000US-0246527.  
 PR 08-NOV-2000; 2000US-0246528.  
 PR 08-NOV-2000; 2000US-0246532.  
 PR 08-NOV-2000; 2000US-0246609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 17-NOV-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.

PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 PI Rosen CA, Barash SC, Ruben SM;  
 XX  
 XX WPI; 2001-465558/50.  
 DR N-PSDB; AAS34879.  
 XX  
 XX Novel polypeptides and polynucleotides useful as diagnostic reagents to  
 PT diagnose diseases or disorders associated with aberrant expression or  
 PT activity of polypeptides, and for treating cancers, Rheumatoid  
 PT arthritis  
 XX  
 PS Claim 11; SEQ ID NO 407; 687bp; English.  
 XX  
 CC The present invention relates to the isolation of novel human neoplastic  
 CC disease associated polypeptides, and cDNA (AAS34767-AAS35050) and DNA  
 CC sequences encoding for these polypeptides. The sequences of the  
 CC invention are useful in the diagnosis, treatment, prevention and/or  
 CC prognosis of disorders involving neoplastic disease such as  
 CC hyperproliferative disorders (e.g. leukemia, bone cancer, bladder  
 CC cancer, brain stem glioma, adult liver cancer, childhood cerebellar  
 CC astrocytoma, or Hodgkin's lymphoma). The sequences of the invention may  
 CC also be useful for treating other disorders such as neural disorders,  
 CC immune system disorders, muscular disorders, reproductive disorders,  
 CC gastrointestinal disorders, pulmonary disorders, cardiovascular disorders  
 CC and renal disorders. The polynucleotide sequences of the invention are  
 CC also useful in gene therapy. AU21568-AU21851 represent the novel human  
 CC neoplastic disease associated polypeptides of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 XX  
 SQ Sequence 203 AA;

Query Match 12.0%; Score 945; DB 22; Length 203;  
 Best Local Similarity 93.8%; Pred. No. 3.3e-56;  
 Matches 180; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

QY 681 PKMGRRSIESPNLGFCTDVILPHLEDDLGSLSEPEPVQNMQRVGVANLTQRE 740  
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 5 pkmgrsiespslgtfctdvlphlleddlgslsepepdqnghtvgkdvagltqre 64  
 QY 741 IDROEVINLEFVASHLRTLVADLIFYOMRKENIMPREELARLPNPETLEINHSW 800  
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 65 idrgevinelfvtashlrtlrvldllyqrmkenimpreelarlpnpelieinhs 124  
 QY 801 CEAMKRLREEGPIRIDISDPMIARFDGPARREILOOVAOFCSYQSVALELRTKQRESR 860  
 |||||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||  
 Db 125 ceamkrlreegpikreisdlmarlfdgpareelqvaafcsysialelktkqrksr 184  
 QY 861 FQLFMQEAESH 872  
 |||||||:|||||  
 Db 185 fqlfmgeaeshp 196

RESULT 14  
 ID ABB62745 standard; Protein; 2559 AA.  
 XX  
 AC ABB62745;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Drosophila melanogaster polypeptide SEQ ID NO 15027.  
 XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
 KW pharmaceutical.  
 XX  
 OS Drosophila melanogaster.  
 XX  
 PN WO200171042-A2.  
 XX  
 PD 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US09231.  
 XX  
 PR 23-MAR-2000; 2000US-191637P.  
 PR 11-JUL-2000; 2000US-0614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 PI Venter JC, Adams M, LI FMD, Myers EM;  
 DR WPI; 2001-656860/75.  
 XX N-PSDB; ABL06848.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions -  
 XX  
 PS Disclosure; SEQ ID NO 15027; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins  
 CC (AB157737-AB172072).  
 CC The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pcl\_sequences.  
 CC  
 XX  
 SQ Sequence 2559 AA:  
 11.9%; Score 936.5; DB 22; Length 2559;  
 Best Local Similarity 20.8%; Pred. No. 5e-54;  
 Matches 481; Conservative 279; Mismatches 652; Indels 895; Gaps 82;  
 QY 17 QSHLSPIASWLSLSLSDSTPRTSPSHRQPSDSTETAGLVORC-----VII 67  
 DB 204 qsdnspvliqapgerslntlptsrldlsghtqesttpattpstlajpknftlly 263  
 QY 68 OKQOHFGFTVSGDRIVLVOSVRPGAAMKAGVKEGDRITKVGWVTNSHLEVKLK 127  
 DB 264 rksnsgymkvsqdpvfvessvpgaaaelaglvagdmllrvnghevrlekhtvvgllk 323  
 QY 128 SGAVALLTLTGSS-----PSVGV-----SGIAQNPSVAGLVRY----- 161  
 DB 324 asttvelavkrqgklrtprsvsvtptslrgtrdtastltpqpvdskkrremetykiq 383  
 QY 162 -----NP----- 163  
 DB 384 tlgkmlgekelnlrksdqdnpsyklseanlrklreqlhgygaedaprvklqaaagknk 443  
 QY 164 --TIIP-----PP-----PP----- 170  
 DB 444 talltpnqiqhlsasachnqgfhllhhnhlnnnnyppqgqastpafsliprslss 503  
 QY 171 ----- 170  
 DB 504 lslgttknktekdltspsfgltctdfliqgrmsghaagsmghstpsqgfthpq 563  
 QY 171 -----PPPIP-----PPO-----HTTG 182  
 DB 564 qghrfketgtakgnkflisrslleedvpprpiprpnprqlnldkngnaaspgshlva 623

QY 183 P-----KPLQDPEVOKHATQILMNN 202  
 DB 624 pvsdlratrspqlnrnsgqqlpssldnspnaskrsklktkalsp-----kmtgmflgm 680  
 QY 203 ----- 202  
 DB 681 esasaagaagslvdgppplpprlpgmmedmarscgqlagpnsvtafaylvstt 740  
 QY 203 -----LROEPEELOIL-----PCGERSOR----- 223  
 DB 741 tavqndnlntafplsqrpnlvqqlqyqgqqlqmsggatgaltgtpnlkxkhrvrs 800  
 QY 224 -----TCEGRLSVDSQADSGLDSTGTEREPSISESLM-----NRNS 259  
 DB 801 spdmhprhpdrtlcttsgsweivekdgess-ppgtpppylsssmvntvledpennrga 859  
 QY 260 VLSDPG--LDSPQ-----TSPVILARVQH--HRRQGSNALPLPHQGIQDSPPRLII 309  
 DB 860 aaagpvgfleshqfcpmagassplislnshmaaqnd-----tqkellis 906  
 QY 310 GPREDYD-----PGYFNNESDIIFODLEKLKSHPAYLVWRYILSQADPGPLFYU 361  
 DB 907 medensdldepfidengpfnlrlleae-----nvflaflnyvinsdpapllfyl 960  
 QY 362 CSEVYQOTNPKDSRLGKDIWNTFLKKNAPLR-VKIPEMIQATIDRLRNEDP-----RN 416  
 DB 961 ltelkyegskdmrkwayelhsflvraplswyrqdeslarevdhvlqleykvelilt 1020  
 QY 417 VLCEAEAVWLLEOEDINDRSKRTGLGSLYEND--LLGLDGDPLREKOMEKOL--- 471  
 DB 1021 vflsrkrakdlisegrlrefqgkrtaglgltyprtdkllaektdlre-qldkylmpn 1079  
 QY 472 -----AALDILSKYEEDRSAPDFA--VNTFMSH----- 499  
 DB 1080 lhaliedengspedrvkvalcesaltvylrntltpsslverhvfsvrsksfksrl 1139  
 QY 500 -----AGIR--LRE-----SRSSCTAKTKQS 518  
 DB 1140 mgknrkmnvrgpnlvltgyevhcnhcgllwvsgpyghctcklnlhrqskvdes 1199  
 QY 519 APDKDKWLPFFPKTKKSSSKKEK-----DALEDKR- 551  
 DB 1200 cggp-----lpqaklalnmdkskfmgtlrptsavigneksrgrdeeldvcltpdrq 1253  
 QY 552 -----NPILRYTGKPKSSSQIKPGVNRNIIQHFNESHQYDVPEPG----- 592  
 DB 1254 asivrgpsdtrpdaaisrsgntscntegll--ntdlqssfhgscandslppggagc 1310  
 QY 593 TORLST-----GSPFE-----DLESDSR-----SEIRLGRS 620  
 DB 1311 mmdltsvastlptstgsvaaglsataelnaidvtvckearterysghpkhkaapvsnrs 1370  
 QY 621 GSLGREEMKRSRKAEN-----VPRPSDVMDMAAARLHGASASSASLS----- 668  
 DB 1371 esykerlnskrnrsrictsdpslsarp-ndegldlglnaelyvgssnslsaagseap 1429  
 QY 669 TRSLENPTPP-----FTFKMGRSIESPNL-----GFCIDVILPHLEDDLGLS--- 713  
 DB 1430 stsmehfaapgaagvqvppmglnqgnhpllllqghaagyccqdsfqaqlaagaagssas 1489  
 QY 714 -----DLEPEEVO--NMOHTVCKDVANVANTOREIDROEYINELFVT 753  
 DB 1490 nssfwmaghplpyarvtlesededvneadwsmaaevlaaltdcekkrrgeilneliyct 1549  
 QY 754 EASHLRTLRLVLDLIFYQRKRKENLMPREIARLF-PNLPELIEIHNSWCEAMKKLR-EEG 811  
 DB 1550 ernhvrctklldrlflplyesgllsqdhlillfpallslstrelhgaefgsqkqrrlenn 1609  
 QY 812 PIIRIDISPMALARFDGPABEELQOVAQFCYSQVALERIKRKOKRESRFQLPOMQAESEH 871  
 DB 1610 hvntitgdlldamfdqsgsvvlcefaqfcargqalealkerknkdmlqkllkxseeh 1669

QY	872	POCRRLDRLDIVSEMQRTRKYPVLLLENIKKH-----EGSTENHEKLCARQDCIELKF	927
Db	1670	kacrrlrlkdllrpvqlrkylllfenlykvtrvlrpenteaaeiqravesskrllve	17299
QY	928	VNEAVKOTENHRLREGYOKRLDATALERASNPLAAEFKSLDITTRKMIHEGPIWIRISKD	987
Db	1730	vngvctaeedhkrqnlqrldrtssydk-----eefkrlldltqphllldhgnlt--lkn	17801
QY	968	KTLDLYVLLLEDVLLVLLQROEERLLK-CHSKTAVGSSDSKOTFSPVYKLNLA-VLIRSVA	1045
Db	1782	psvqjhgjlifemimvllttkqdkkyllknlhtrpisi-----tnkxpspimsidatdlirgea	1837
QY	1046	TDKRAAFITCSLEIGPQPIQYELVALNSSDKNIMELLEEVAONAKH-----	10922
Db	1838	adksnfll---kmltqgmlelrlrpssecktwfkhlfsdvaarqsknrsknaashndtsi	1894
QY	1093	---PCAAPPIHPSPSGOE-----	1118
Db	1895	sdpralaarp-hsnkleslstdtvgplaatallttplaplmlplatvtrpapatnmsvs	19533
QY	1119	-----RVEINDESVNTEKERKKLPREGGPORVQDKOLI-AOGEPPVO-----BED	11633
Db	1954	sltvgvrlngprdataesadadyvnpkprssqnevrnlrmtsrstsegrpqkysangtean	20133
QY	1164	EEELR-----TLPRAPSLDGENR-----GIRTDPLVAL-----TGPLMEG-LA	1204
Db	2014	dvlrlrhsgstresvtrpsltgeenstcymvgwnskrdsasivcsnmsntrllmqspdv	20733
QY	1205	D-PALEDEVENRHLILMSLPGHTVKVQAAGEBED-----DLTPRP-----SVYSI	1249
Db	2074	dptaiqvispahhaepvlrpgcklrrldasindlilekqkllcdifrlrphenyqdvld	21333
QY	1250	TSHPWDPSGPQAFITISDTRLARPEGSQ-----PEGEDVAVSSL-----AH-	1291
Db	2134	ammp-----eapkdadaialaaydqifqltkmlneymhvrpegevaavstavgchhe	2186
QY	1292	-----LPPRTSSGVWMDSPELDNRPAAEASTPPAASYVVRKVSILPEGGYGAAKV	13433
Db	2187	kekllrkvpasstfsspprlpprrqhagaqqlpps---rlmpklqltlddev-----	2238
QY	1344	AGSNAIDDSGGSESELEVEVGAGQANGCNCYVSMRPGPLDSSTPFPMPGPPSPSOCHSLPA	14033
Db	2239	---alhedddgyceidelt-----lpaispkpherpl-cplap-----	2272
QY	1404	WPTPEPQRYRGVGGQCSILVRDNDVIFHTI-----EQLTIKLRL-	1444
Db	2273	ntepktsqsv-----laskrtgstdavpeglllegepdlgetkgetgednevktvpskils	2328
QY	1445	---KDMELAHRELKLSIGESSGGTTPVGSFTEAARTDYSLSPPAKALASDSQNGQ	1500
Db	2329	eschneeqcvceadttkkadvaplttskneaaasvdlpsgrtelktaensksvadckehne	2388
QY	1501	EGSGPREGSDIALEDSATDTAVSPGP	1527
Db	2389	e---tleegy-astlvdsstlqtsptseap	2411
RESULT 15			
AAM39775			
ID	AAM39775	standard; Protein; 696 AA.	
XX	AAM39775;		
XX	22-Oct-2001	(first entry)	
DE	Human polypeptide SEQ ID NO 2920.		
KW	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	anotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukaemia.		

[illegible]

```
Db 240 ryaltnkalifvismdng-aqiyelvaqvsekltwqdlcr-maasvkegstxpiip 297
QY 1102 PSPPG-----SOEPAYOGSTSSRVEI 1122
Db 298 gsteprgegdndeedprklikeeghysvtglsqspdrdlglestlisskpgshsistsgksey 357
QY 1123 ND-----SEVYHTEKEEPKILPEG-----PBPBQR-VOD-----KOLIAQ 1155
Db 358 rdlfvaergfakeghdgtclikevgedyqlapshlpvseerwaldalrnlglkqllvq 417
QY 1156 GEPVOED-EEELRTLPRAPRPSLDGENRGI RTDPVYLLATGFLMEGLADALEDEVNL 1214
Db 418 qlgltexsvgedwqhfr-----yrt-----asgpr-----qtdsviqnseni 455
QY 1215 RHLIWSLLPGHTVKTQAGEEPEDLTPTPSVVSIHSHWDPGSPGOAPTISDSTRL--- 1271
Db 456 k---ayhsgeghmpfrftgtgdlatcysprtsstefa--prd--svglapqdsqasnllvm 508
QY 1272 -----ARPEGs-OPEGEDVAVSSLAHLPPR-TRSSGVWDSPELDRN----- 1310
Db 509 dhmlmtpemprmepeglddsgehfddareahsdenpsgdqavnkkekdvnlrlisgnyl 568
QY 1311 -----PAAEAASTEPAAGYKVVRKVSLPGCGVGAAYAGSNA-IPDSGQSESELS--- 1360
Db 569 lldgydpvgesscdeevas-----siltqpmctgipavesltghqghspqnthsgaalspft 623
QY 1361 -----EVEGGAQATNCFTVVSMPAGPLDSSTEPTGTTPSPSQCHSLPAMPTEPQPYRGVR 1415
Db 624 peflvqgrwgamey-scfeiqspscadsqsg----- 654
QY 1416 GGCSSLVRRDVVIFHTIEQLTIKHLRLKDMELAHRELKXSGES 1462
Db 655 -----lmeylhkleadlhlkveesytlcqrlagsa 687
```

Search completed: September 18, 2002, 10:37:54  
Job time: 116 sec